Optimising Tests of Primate Cognition:
Towards Valid Species Comparisons

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Michèle N. Schubiger

von

Gommiswald SG

Promotionskommission

PD Dr. Judith M. Burkart (Vorsitz & Leitung)

Prof. Dr. Carel P. van Schaik

Prof. Dr. Marta B. Manser

Zürich, 2019

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

I ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to my interdisciplinary dyad of supervisors, PD Dr. Judith Burkart and Prof. Carel van Schaik, who guided me through this PhD journey and were both always available for spontaneous meetings to exchange ideas. I am particularly grateful to Judith, my main supervisor, for providing me with quick advice during the cognitive testing phases and with helpful feedback on manuscripts during the writing phases. I am also very grateful to Carel, my second supervisor, for his patience and ongoing scientific enthusiasm and for sharing his multi-faceted thoughts as the focus of my thesis began to crystallise. Our discussions broadened my perspective on primate cognition and the evolution of non-human and human primates in general.

Furthermore, I thank Prof. Marta Manser for being part of my thesis committee, and Prof. Thomas Bugnyar for agreeing to review this thesis.

During the experimental phase of my PhD, I very much appreciated the help of several dedicated research assistants and collaborators. I would like to thank them in the order we met and worked together: Chantal von Siebenthal (marmosets) and Piero Amodio (squirrel monkeys) for assisting with habituation of the monkeys to the new test enclosures, Dr. Florian Wüstholz and Dr. André Wunder for introducing their philosophical approach to definitions and research questions to the Primate Station and assisting with video coding, and Alexandra Kissling for her help with experimental data collection.

A special thank you goes to the whole Primate Station team, including all the non-human primates it was such a pleasure to work with. My studies would not have been possible without the dedication and flexibility of our animal care team. I am grateful to Heinz Galli, for help with building my experimental apparatuses, and to Dr. Claudia Rudolf von Rohr, Thomas Bischof, and Patricia Rivera for always being interested and helpful and taking such great care of the monkeys. Furthermore, I thank Zaida Kosonen for introducing me to the squirrel monkeys and the behavioural coding software, and Ruben Richiger for his help with converting videos. My warmest thanks go to Elo(ísa) Martins for providing me with one of her great marmoset photographs for my publications and for sharing part of this journey with me. Saudades! Last but not least, I am immensely grateful to Dr. Christa Finkenwirth, my long-term office mate, for being so direct, honest and open-minded, for sharing all our ups and downs along the way, and for ensuring that I was never short of oxytocin.

At the Department of Anthropology, I am thankful to PD Dr. Karin Isler for passionately and patiently sharing her statistical expertise and to Dr. Erik Willems who also provided me with statistical advice. A big thank you goes to the department's administration team who were enormously supportive and friendly. Claudia Zebib and Ruth Haegi continuously ensured all administrative processes ran smoothly. Marcus Gisi resolved several computer issues such as the quality of my graphs and ensured my numerous videos were safely stored on the server. Dr. Tony Weingrill was responsible for the Evolutionary Biology program, including the very interesting and fun PhD retreats from Lake Lugano to the alps of Valois.

I was fortunate to meet many collegial and enthusiastic people from several interesting research fields within and outside the department. Besides all Anthropology people, I would like to thank the members of the Coop-Peer-Action group for several years of successful activities, exchanging experiences in research and life, and motivating each other during our yearly writing retreats. I also thank the CPA ladies and the 'external' scientific advisors, Professors Barbara König (IEU), Marta Manser (IEU), and Redouan Bshary (University of Neuchâtel) for valuable feedback on my project during several research workshops.

Finally, I would like to thank my parents, sister and my brother and his family for their love and support and my grandmothers whom I keep in loving memory. Invaluable has been the emotional support of my friends during those years. In particular, I would like to extend my warmest thanks to Nicole for always being there for mindful discussions about work and life with a healthy pinch of humour, and to Ulli for still being there after nearly three decades of friendship.

Most of all, I thank Nic for sharing the same passion, for always being there for me despite all those years of not even being able to live on the same continent or in the same country. Nic, your creativity and positivity are an inspiration and words cannot express how grateful I am to have you in my life.

II THESIS ABSTRACT

Comparative cognition research aims at assessing and comparing the cognitive abilities of non-human primates and other non-human animals, typically in order to link variation in cognitive abilities to variation in brain size or socio-ecological factors. Ultimately, this allows us to enhance our understanding of how the human mind works and how our mental processes might have evolved. While such cognitive comparisons have traditionally been conducted between different species and have largely ignored within-species variation in cognitive abilities, recent research efforts are beginning to focus on such individual-level differences. Understanding to what degree individuals of a single species differ in cognitive abilities is vital to be able to reliably infer species differences. This fundamentally includes to determine how much of the inter-individual variation can be explained by cognitive ability per se rather than non-cognitive factors. Only then will we be able to improve existing cognitive tests and develop new ones that reliably measure cognitive performance and allow for fair species comparisons.

This thesis aims at identifying and quantifying the influence of such non-cognitive factors that may affect cognitive performance. It contributes to a recent line of research that has begun to assess such potentially influential factors in an attempt to optimise the measurement validity and reliability of primate cognition tests and ultimately allow for fair comparisons of cognitive abilities both within and across species.

In a series of three experimental studies (*Chapters 2-4*), I investigated whether individual differences in a number of non-cognitive factors lead to individual differences in cognitive performance in physical cognition tasks and whether improving test designs helps both the internal (measurement accuracy) and external validity (generalisability) of cognitive tests.

In my first study (*Chapter 2*), I investigated if individual differences in emotional reactivity lead to individual differences in cognitive performance. I addressed in common marmosets (*Callithrix jacchus*), whether those individuals who tend to be highly emotionally aroused during cognitive testing perform differently in an object permanence task compared to their conspecifics who remain calm. My findings revealed substantial individual variation in the marmosets' emotional reactivity that affected their motivation to participate.

Marmosets, mainly males, who showed a strong spontaneous emotional reaction to the experimenter and the test situation, were often distracted and participated in fewer trials. Importantly, however, their cognitive performance in the object permanence task remained unaffected. Highly emotionally aroused males performed equally well in those test trials in which they did participate. Importantly, however, this was only the case when I controlled performance for attention and because I used clear pre-defined criteria that determined when to stop a test session prematurely owing to lack of motivation. Therefore, when testing non-human primates with cognitive tasks researchers need to be aware of inter-individual differences in the subjects' emotional reactivity and their potential effects on participation and attention. Ideally, attention to the test stimuli is quantified and taken into account when analysing the data, and clear stop criteria are consistently applied to avoid testing subjects who are not fully motivated to engage with the test. In such cases, a test session should be discontinued and only resumed later when the subject is more motivated to participate.

In my second study (*Chapter 3*), I investigated whether task format, i.e., how the cognitive test is designed and the test stimuli are presented to the subject, affects cognitive performance. Specifically, I focussed on one widely used aspect of task format, i.e., the number of test stimuli in typical choice paradigms in which a subject has to make a correct choice among several options in order to pass a test trial. I found that when I tested the marmosets with a classical two-choice memory test in which they had to remember the location of a food item, their performance did not decrease with increasing memory demands (i.e., longer time delays). This indicated that this memory test did not measure memory performance and its validity needed improvement. To determine if the subjects had not been sufficiently motivated to choose correctly because even choosing at random would give them a 50:50 chance to discover the food reward, I tested a second sample of marmosets and squirrel monkeys (*Saimiri sciureus*) with a novel nine-choice memory test version and the two-choice version. My results confirmed that these experimentally naïve monkeys performed equally poor in the two-choice version. In the nine-choice memory test, however, they performed significantly better and showed longer retention ability, and their performance decreased with increasing memory demands as predicted. My findings thus highlight that the very frequently used two-choice paradigm is problematic to quantify cognitive abilities in marmosets and squirrel monkeys, and most likely in other species too.

However, small modifications to the test design can significantly increase the measurement validity of such cognitive tests.

In my third study (*Chapter 4*), I explored if individual differences in the amount of required testing time lead to individual differences in cognitive performance. Some subjects are less consistently motivated to participate in a full test session per day and require more time to complete the task. Since access to non-human primates is often limited, researchers often have to exclude these subjects from testing and, if possible, replace them with other subjects who are more consistently motivated to participate. However, such opportunistic testing could bias the results if these dropouts differ systematically in cognitive ability. In fact, it is well conceivable that some individuals are less interested in participating because the task is more difficult for them compared to their more motivated conspecifics. To explore this possibility, I tested marmosets and squirrel monkeys with a test battery consisting of a range of tasks from several physical cognition domains. Importantly, in our full testing approach, we minimised dropouts by giving each subject sufficient time to complete the tasks, even if this would take a considerable number of test sessions. A direct comparison of subjects who completed the task within a typical time limit (and would have been included in opportunistic testing) and subjects who completed the tasks after the time limit would have expired (and would have dropped out) revealed no difference in cognitive performance. My findings show that opportunistic testing does not necessarily bias the results of cognitive studies. Although full testing is often important, such as in general intelligence testing where the same subjects have to be tested with a whole test battery and dropouts critically affect statistical analysis, opportunistic testing may be a valid alternative in other studies when full testing is not feasible.

My thesis contributes to the recent paradigm shift in comparative cognition and related research disciplines that acknowledge and focus on individual-level differences in cognitive abilities. Furthermore, my findings add to the recently growing body of empirical literature on non-cognitive factors whose potential influence on test performance have to be taken into consideration in order to optimise existing and develop novel tests of primate cognition. Once this task is achieved, we will be able to conduct fairer and more meaningful species comparisons in order to better understand how much of our evolutionary heritage we share with other primates and how the exceptionally complex human mind may have evolved.

III ZUSAMMENFASSUNG

Die vergleichende Kognitionswissenschaft erforscht die intellektuellen Fähigkeiten nicht-menschlicher Primaten und anderer Tiere, üblicherweise um einen Zusammenhang zwischen der Variation dieser Fähigkeiten und der Variation von Hirngröße oder sozio-ökologischen Faktoren herzustellen. So erweitert sie letztlich unser Verständnis davon, wie der menschliche Intellekt funktioniert und wie unsere mentalen Prozesse im Laufe der Evolution entstanden sein könnten. Traditionelle Studien verglichen bis vor kurzem in erster Linie die kognitiven Fähigkeiten verschiedener Arten. Unterschiede innerhalb einer Art wurden weitgehend vernachlässigt. Neuere Studien haben jedoch damit begonnen, den Fokus auf Unterschiede zwischen Individuen innerhalb einer Spezies zu verlagern. Ein besseres Verständnis individueller Unterschiede innerhalb einer Art bildet eine wesentliche Voraussetzung, um Rückschlüsse auf Artunterschiede ziehen zu können. Es gilt insbesondere herauszufinden, welcher Anteil der inter-individuellen Variation sich tatsächlich durch kognitive Fähigkeiten und nicht durch andere Faktoren erklären lässt. Dann erst werden wir in der Lage sein, bestehende kognitive Tests zu verbessern und neue Tests zu entwickeln, welche die kognitive Leistung zuverlässig messen und faire Artenvergleiche ermöglichen.

Ziel dieser Dissertation ist es, nicht-kognitive Faktoren, welche die kognitive Leistung beeinflussen können, ausfindig und quantifizierbar zu machen. Damit liefert sie einen Beitrag zu einer neuen Forschungslinie, die bereits mit der Evaluierung solch potenzieller Störfaktoren begonnen hat, um basierend darauf die Validität und Reliabilität kognitiver Tests für Primaten zu optimieren. Dies soll schliesslich faire Vergleiche kognitiver Fähigkeiten sowohl innerhalb als auch zwischen Arten ermöglichen.

In drei experimentellen Studien (*Kapitel 2-4*) erforschte ich mittels physikalischer Kognitionstests, ob individuelle Unterschiede im Bereich nicht-kognitiver Faktoren zu individuellen Unterschieden in der Testleistung führen und ob Verbesserungen der Testdesigns die interne (Messgenauigkeit) und externe Validität (Generalisierbarkeit) kognitiver Tests erhöhen.

In meiner ersten Studie (*Kapitel 2*) untersuchte ich, ob individuelle Unterschiede in der emotionalen Reaktivität zu individuellen Unterschieden in der kognitiven Leistung führen. Dabei ging ich der Frage nach, ob Individuen, die während kognitiver Testungen zu

erhöhter emotionaler Erregung (Arousal) neigen, anders abschneiden als ihre gelasseneren Artgenossen. Meine Ergebnisse zeigten beträchtliche individuelle Unterschiede in der emotionalen Reaktivität von Krallenaffen, die sich auf ihre Motivation, an einem Objektpermanenz-Test teilzunehmen, auswirkten. Dies traf vor allem auf männliche Krallenaffen zu. Sie reagierten emotional sehr stark auf den Experimentator und die Testsituation, waren oft abgelenkt und nahmen an weniger Testdurchgängen teil. Entscheidend war jedoch, dass ihre kognitive Leistung nicht von der Ausprägung ihrer emotionalen Reaktion betroffen war. Emotional hochreaktive Männchen schnitten in jenen Testdurchgängen, an denen sie teilnahmen, gleich gut ab wie ihre Artgenossen. Zu diesem Ergebnis kam ich, wenn ich nur die Leistung in jenen Testdurchgängen berücksichtigte, in denen die Affen aufmerksam waren. Entscheidend war zudem, dass ich klare, vordefinierte Abbruchkriterien verwendete, die bestimmten, wann eine Testreihe wegen mangelnder Motivation vorzeitig zu beenden war. Beim Testen kognitiver Fähigkeiten nicht-menschlicher Primaten müssen sich Wissenschaftler folglich solcher individueller emotionaler Reaktivitätsunterschiede und derer möglichen Auswirkungen auf Teilnahme und Aufmerksamkeit der einzelnen Tiere bewusst sein. Idealerweise sollte die Aufmerksamkeit gegenüber den Teststimuli quantifiziert und bei der Datenauswertung berücksichtigt werden. Ebenso sollten klar definierte und einheitliche Abbruchkriterien verhindern, dass Primaten getestet werden, die nicht gänzlich motiviert sind. In solchen Fällen ist es ratsam, die begonnene Testsitzung zu beenden und erst dann fortzusetzen, wenn ein Individuum wieder motiviert ist, teilzunehmen.

In meiner zweiten Studie (*Kapitel 3*) untersuchte ich, ob das Aufgabenformat, d. h. die Art und Weise wie ein Test konstruiert ist und wie die Teststimuli dem Affen präsentiert werden, die kognitive Leistung beeinflusst. Ich konzentrierte mich dabei auf ein weit verbreitetes Testformat, bei dem der getestete Primat aus mehreren Optionen eine korrekte Wahl treffen muss, um einen Testdurchgang zu bestehen. Zunächst konfrontierte ich Krallenaffen mit einer klassischen Gedächtnisaufgabe mit zwei Auswahlmöglichkeiten: Sie mussten sich über ansteigende Zeitintervalle merken, wo sich ein Futterstück befand. Ihre Leistung nahm jedoch mit steigenden Gedächtnisanforderungen, d. h. längeren Verzugsintervallen, nicht ab. Dies wies darauf hin, dass der verwendete Gedächtnistest die Erinnerungsleistung der Krallenaffen nicht zuverlässig messen konnte und dessen Validität verbessert werden musste. Ich wollte herausfinden, ob die Affen schlichtweg nicht

hinreichend motiviert waren, das korrekte Versteck zu wählen, weil sie sogar bei einer rein zufälligen Wahl eine 50:50-Chance hatten, die Futterbelohnung aufzudecken. Daher konfrontierte ich eine zweite Gruppe von Krallenaffen und Totenkopfäffchen sowohl mit einer neuartigen Gedächtnis-Aufgabe mit neun Auswahlmöglichkeiten als auch mit der herkömmlichen Version mit nur zwei Optionen. Meine Ergebnisse bestätigten, dass diese Individuen, die vorher noch nicht getestet worden waren, in der letzteren Testversion ebenso schlecht abschnitten wie ihre Vorgänger. Im Gedächtnistest mit neun Wahlmöglichkeiten hingegen schnitten beide Spezies signifikant besser ab und ihre kognitive Leistung nahm nun, wie erwartet, mit steigenden Anforderungen an die Erinnerungsfähigkeit ab. Meine Ergebnisse veranschaulichen, dass sich das häufig eingesetzte Testparadigma mit zwei Wahlmöglichkeiten zur Quantifizierung der kognitiven Fähigkeiten von Krallenaffen und Totenkopfäffchen, und sehr wahrscheinlich anderer Arten, nicht eignet. Jedoch können bereits kleine Veränderungen des Testdesigns die Messgültigkeit solcher kognitiven Tests erhöhen.

In meiner dritten Studie (*Kapitel 4*) erforschte ich, ob individuelle Unterschiede in der benötigten Testungszeit zu individuellen Unterschieden in der kognitiven Leistung führen. Manche Individuen sind weniger anhaltend motiviert, an einer vollen Testsitzung pro Tag teilzunehmen. Sie benötigen mehr Zeit, um eine Testaufgabe abzuschließen. Da der Zugang zu nicht-menschlichen Primaten oft eingeschränkt ist, müssen Forscher solche Individuen oft von der Testung ausschliessen und durch andere ersetzen, die kontinuierlicher an der Teilnahme motiviert sind. Solch opportunistisches Testen könnte jedoch dann die Resultate verfälschen, wenn sich die Testabbrecher (Dropouts) hinsichtlich ihrer kognitiven Fähigkeiten systematisch unterscheiden. Es ist durchaus denkbar, dass diese Individuen deshalb weniger interessiert sind, weil der Test für sie schwieriger ist als für ihre motivierten Artgenossen. Um diese Möglichkeit zu prüfen, testete ich Krallenaffen und Totenkopfäffchen mit einer Testbatterie, die sich aus einer Reihe von Tests aus mehreren physikalischen Kognitionsdomänen zusammensetzte. Bei diesem umfassenden Testungsansatz war entscheidend, dass ich die Dropouts auf ein Minimum reduzierte, indem ich jedem Teilnehmer ausreichend Zeit zur Verfügung stellte, die Testaufgaben zu bewältigen. Ein direkter Vergleich jener Individuen, die eine Aufgabe innerhalb eines typischen Zeitlimits abschlossen (und bei opportunistischem Testen in der Stichprobe verblieben wären) mit jenen, die sie erst abschlossen, nachdem das Zeitlimit abgelaufen war (und daher

ausgeschlossen worden wären) ergab keinen kognitiven Leistungsunterschied. Meine Ergebnisse zeigen daher, dass opportunistisches Testen nicht zwangsläufig zu einer Verfälschung der Resultate kognitiver Studien führt. Vollumfängliches Testen ist oft wichtig, wie etwa bei allgemeinen Intelligenztests, bei denen dieselben Individuen mit einer ganzen Testbatterie getestet werden müssen und Dropouts die statistischen Analysen entscheidend beeinträchtigen. In anderen Studien kann opportunistisches Testen jedoch eine valide Alternative sein, wenn eine vollumfängliche Testung nicht umsetzbar ist.

Meine Dissertation trägt zum kürzlich erfolgten Paradigmenwechsel in der Vergleichenden Kognitionswissenschaft und verwandten Disziplinen bei, der individuelle Unterschiede kognitiver Fähigkeiten anerkennt und diese zum Forschungsgegenstand macht. Darüber hinaus erweitern meine Erkenntnisse den neuerdings wachsenden Fundus an empirischer Literatur zu non-kognitiven Faktoren, deren möglicher Einfluss auf die Testleistung in Erwägung gezogen werden muss, um bestehende kognitive Tests für Primaten zu optimieren und neue Tests zu entwickeln. Erst wenn uns dies gelungen ist, werden wir in der Lage sein, fairere und aussagefähigere Artenvergleiche durchzuführen und besser zu verstehen, wie viel unseres evolutionären Erbes wir mit anderen Primaten teilen und wie sich unser ausserordentlich komplexer menschlicher Intellekt im Laufe der Evolution entwickelt hat.

Chapter 1

General Introduction

1.1 Comparative cognition

1.1.1 What is comparative cognition?

Comparative cognition is an interdisciplinary field of research that focuses on understanding and comparing mental processes in humans and other animals and aims to reconstruct how cognitive abilities evolved by comparing various extant species. One central goal, particularly of psychologists and anthropologists working in this field, is to better understand how the human mind works and evolved. Ultimately, by identifying to what extent we share cognitive abilities with other animals, we aim to reconstruct how our elaborate cognition may have evolved and how we became such a successful species. In order to reconstruct how a given cognitive ability may have evolved, comparative psychologists, for instance, use an experimental task to investigate if the ability or the capacity to acquire the ability, is present in two or more species. Since pre-linguistic human infants and nonhuman animals cannot be asked verbally about what they know about their physical and social environment, their cognitive abilities have to be inferred from their performance in human-designed cognitive tasks.

To test for different evolutionary scenarios, researchers often opt to compare species with a particularly high or low degree of phylogenetic relatedness, and who vary with regard to potentially influential social or ecological factors. If two evolutionary closely related species perform well in the cognitive task, it is inferred that the required mental capacity to succeed was already present in the last common ancestor the two species shared in their evolutionary history. If two evolutionary distantly related species succeed in the cognitive task, whereas the trait is absent in their close relatives, it is concluded that convergent cognitive evolution took place, meaning that the capacity for the underlying mental ability evolved independently and at different times after the split from their last common ancestor. Another possible scenario is that only one of the two species performs well suggesting that different adaptations in response to different physical or social environments play a role. If only species A passes the task, this indicates that the cognitive capacity only evolved in this species but not in species B, presumably because species B was not exposed to similar challenges and selection pressures (Boyd & Silk, 2012). A critical precondition to justify such often far reaching conclusions is, however, that different species are tested in a fair manner (see *1.1.2* to *1.1.4*) and that a cognitive task truly measures the cognitive ability it was

designed to measure, a challenge we should not underestimate and to which I will turn later (see *1.2.3ff*).

1.1.2 Is it possible to compare cognition between different animal species?

Conducting fair comparisons of cognitive abilities across a wide range of different animal species turns out to be a considerable challenge for the field of comparative cognition. In fact, it is such a challenging mission that some researchers have argued that it will remain an impossible one (e.g., Bitterman, 1965, 1988; MacPhail, 1982, 1987). Bitterman for instance, argued that comparing the cognitive abilities of vertebrate and invertebrate species in a fair manner will never be possible because of profound differences in their anatomy and perception. Intriguingly, some invertebrates, such as octopuses and other cephalopods, possess a highly developed visual system, high levels of dexterity and have shown remarkable cognitive abilities (Mather & Kuba, 2013; Tricarico et al., 2014) indicating some degree of convergent evolution of the distantly related invertebrates (particularly the mentioned molluscs) and vertebrates (e.g., some mammalian and some bird species). However, despite these fascinating analogies, Bitterman may be correct that biological (sub)phyla are perhaps not the best starting point for fair cognitive comparisons. Even when narrowing down such comparisons to vertebrates, different taxa still differ substantially in a range of anatomical (e.g., dexterity) and perceptual (e.g., vision) features and these physical differences in turn may mask or exaggerate differences in their mental abilities. For instance, some bird species, particularly corvids and parrots, possess particularly large brains in comparison to other birds (Cnotka et al., 2008; Iwaniuk et al., 2005; Sultan, 2005). Moreover, some of the structures in avian brains, particularly the neopallium in the forebrain, have been classified as homologous to the mammalian neocortex which plays a central role in higher cognitive abilities (Güntürkün & Bugnyar, 2016; Jarvis et al., 2005). Corvids and parrots have indeed demonstrated remarkable cognitive performance in several experimental paradigms (e.g., tool manufacture and use: New Caledonian crows, *Corvus moneduloides*, (Hunt, 1996; Hunt & Grey, 2004), episodic-like memory: Florida scrub jays, *Aphelocoma coeruleus*, (Clayton & Dickinson, 1998); and mirror self-recognition in magpies, *Pica pica* (Prior et al., 2008)). Some of these findings suggest convergent evolution of cognitive abilities between birds and primates (e.g., Emery & Clayton, 2004) and some researchers have claimed that

some of these cognitive abilities equal those of great apes (e.g., Emery, 2006). However, despite their highly developed brains and accurate vision, interacting with the same experimental apparatuses that are used for primates in order to solve a cognitive problem may prove difficult since birds rely almost exclusively on using their beaks. Even when narrowing down species-comparisons further to mammals other than primates, similar issues remain. Canids such as wolves, domestic dogs, hyenae, and foxes are relatively small-brained and have relatively poor vision. Moreover, like birds, they do not have hands, a physical constraint that has to be taken into account when adapting cognitive test apparatuses and paradigms that have initially been developed for primates who have the highest levels of dexterity in the animal kingdom (Begun, 2015). The opposite is the case for the canid sense of smell which is so highly developed in comparison to primates (with the exception of some prosimians, Rushmore et al., 2012) that the odour of experimental stimuli, i.e., food rewards, has to be masked in order to prevent canids from solving a task by picking up olfactory cues rather than visual or auditory ones (e.g., Miklósi et al., 2003). Large-brained mammals such as elephants and dolphins also rely on using their noses to handle experimental apparatuses. Thus, valid quantifications of cognitive abilities across species are difficult to achieve, and these difficulties are more prominent the more these species vary from each other morphologically, perceptually, and phylogenetically.

1.1.3 Comparative cognition within primates

As I mentioned earlier, studying and comparing other animals' minds can help us to reconstruct how the human mind evolved and this task is most convincingly achieved by studying the cognitive abilities of our closets living relatives, the non-human primates. Despite the above-mentioned limitations of between-species comparisons, more conservative comparisons should be legitimate, at least within the primate order, even though clear differences in anatomy and perception are also present between different primate taxa. Primates are divided into two suborders, prosimians and anthropoids. Prosimians (lemurs and lorises), who are mostly nocturnal, differ considerably from anthropoid primates in their olfactory and visual systems, which is particularly expressed in an enhanced sense of smell and their reduced ability to perceive and distinguish colours, particularly along the red-green spectrum (Gilad et al., 2004). In order to test nocturnal

species, such as mouse lemurs *(Microcebus murinus*), in an ethical and fair manner, researchers test them under infrared light (Kittler et al., 2018). Even within anthropoid primates that include the tarsiers, New World and Old World monkeys, and small and great apes, remarkable differences in anatomy and perception remain. For instance, one New World monkey family, the aotids (also named owl or night-monkeys), became secondary nocturnal during the course of evolution and both their vision (reduced colour vision) and olfaction (enlarged olfactory bulbs) adapted to this night-active lifestyle, two physiological features they share with prosimians rather than other anthropoids (King, 2016). Even among the diurnal New World monkeys most of whom possess typical anthropoid vision, some morphological differences remain. All members of the callitrichid family (marmosets and tamarins), for instance, have claws rather than typical primate hands with finger nails (giving them the German name "Krallenaffen" which translates to "claw monkeys"), and some individuals, mostly male ones, have reduced colour vision regarding the red-green colour spectrum, similar to prosimians and aotids (Freitag et al., 20012; Pessoa et al., 2005). When testing marmosets, my main study species in this thesis, it is therefore important to build experimental apparatuses that are easy to handle with claws and to only use colours that all subjects can distinguish, i.e., blue and yellow experimental stimuli rather than red and green ones (Schmonn, 2011) or avoid colour altogether by simply using black and white components (Strasser & Burkart, 2012; Schubiger et al., 2015, 2016, 2018). Finally, even among Old World primates and between small and great apes, differences in limb morphology such as how flexible their hands are and whether they have opposable thumbs affect their ability to manipulate objects (Heldstab et al., 2016).

1.1.4 Comparative primate cognition at the species-level

The majority of past studies on the cognitive abilities of non-human primates have focussed on comparing two or more species regarding one or several cognitive abilities (e.g., Amici et al., 2008; Anderson et al., 2010; Mac Lean et al., 2014; Mulcahy et al., 2005; Petit et al., 2015). However, since fair species comparisons are difficult even when restricting them to the primate order and further to primate genera within the two suborders, it appears most sensible to start at the species level and compare individuals of the same species. Individuals of the same primate species may vary considerably in terms of cognitive abilities, but their

cognitive performance may also be influenced by inter-individual differences in non-cognitive factors. In my PhD project, I took this latter and newer approach that allowed me to first identify the factors that influence cognitive performance in a single primate species, and, in a next step, to directly compare them to another species. In order to compare the cognitive abilities of individuals of my main study species, common marmosets (*Callithrix jacchus*) and to create conditions to later compare them to other primate species, it was essential to use cognitive tasks that are solvable regardless of a species' evolutionary adaptations to its ecological niche. Therefore, in contrast to the commonly used approach to use ecologically valid tasks which is valuable in its own right, I focused primarily on tasks used in test batteries to assess general intelligence, an empirical entity that correlates with an individual's ability to solve a diverse range of problems, largely independent of its niche-specific adaptations, and that can only be assessed using reasonably abstract tests.

1.1.5 Domain-generality versus modularity of cognitive abilities

How the human mind evolved is a vividly debated topic that has attracted much interest in the scientific community. Currently, researchers from several disciplines still disagree if or to what extent natural selection acted on the evolution of domains of cognitive abilities to deal with specific problems, on so-called higher cognitive abilities in general that can be applied to a variety of different contexts, or a combination of the two (for an extensive review, see Burkart, Schubiger, & van Schaik, 2017a, b).

Among the proponents of the domain-specific view, there is currently little consensus on what a domain actually constitutes. Depending on the research discipline, cognitive domains have been described from broad clusters of similar cognitive abilities such as a physical versus a social domain (comparative psychologists, e.g., Tomasello & Call, 1998, Herrmann et al., 2007) to highly specialised modules that can be localised in the brain and supposedly adapted to solve a particular problem that our ancestors repeatedly encountered as far back as in the Pleistocene (evolutionary psychologists, e.g., Cosmides & Tooby, 1989, 2002, 2013). The physical domain entails those cognitive abilities that help an individual to understand and adequately respond to problems posed by its physical environment such as gravity and connectivity that help it to find food and avoid predation. Cognitive abilities from the social domain on the other hand help an individual to understand and interact with its

social environment such as finding mates and cooperating or competing with conspecifics. Examples of specialised brain-modules in humans that supposedly evolved for a single purpose are the so-called “cheater detector” and a brain area responsible for the recognition of faces (Cosmides et al., 2005; Ermer et al., 2007; Fiddick et al., 2005; Stone et al., 2002).

The first proponents of the domain-general view (differential psychologists, e.g., Spearman, 1904, 1927) studied stable individual differences across a broad range of cognitive abilities in humans. For this purpose, they developed and validated measuring instruments, so-called general intelligence test batteries, that consisted of numerous subtests, each of which assessing a particular cognitive ability. After more than a century of testing large human samples, these test batteries are now standardised and normed for different populations of people and there is now ample evidence and wide acceptance for (domain)general intelligence in humans (Deary et al., 2010). General intelligence corresponds to the positive manifold, i.e., the fact that overall, all the cognitive tests scores are positively correlated across individuals and result in a single factor ‘g’ in factor-analytical approaches (see *1.2.1*). From an evolutionary point of view, an intriguing question is whether general intelligence is unique to humans or whether there is evolutionary continuity across the animal kingdom. In chapters 1.2 and 1.3, I point out that domain-general cognitive abilities in non-human animals have so far received much less attention which is reflected in the scarcity of suitable measurement instruments. This lack of interest in domain-general cognitive abilities in non-humans is partly caused by the fact that evolutionary scenarios of how cognitive abilities may have evolved seem more straightforward if one assumes that the animal mind is composed of a small number of modules to which other modules are added whenever needed (Clune et al., 2013; Geary & Huffman, 2002; Sherry, 2006).

1.2 General intelligence

1.2.1 What is general intelligence?

In my thesis, I will use Gottfredson’s (1997) broad intelligence definition when speaking of general intelligence. According to Gottfredson, intelligence “is a very general capability that among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather it reflects a broader and

deeper capability for comprehending our surroundings – 'catching on', 'making sense' of things, or 'figuring out' what to do." When speaking of non-human primates, I will use a simplified definition of general intelligence, namely reasoning ability and behavioural flexibility.

Psychometrically, general intelligence correlates with the general intelligence factor *g* that emerges in humans when their test scores from a cognitive test battery consisting of a wide range of different tasks are analysed using factor-analytical statistical methods: a person who performs well in one cognitive test will likely do so in many others. This *g*-factor, the positive correlation of an individual's cognitive performance scores across domains, is also known as the 'positive manifold'. In humans, the existence of *g* is now well established and widely accepted (Carroll, 1993; Deary et al., 2010). In non-human primates, however, far less effort has been invested into developing cognitive tests to investigate whether similar consistent individual differences exist within a single species.

Meta-analyses across many non-human primate species delivered some evidence for domain-generality of cognition, i.e., they found that between-species variation in performance across various cognitive tasks is best explained by a single factor '*G*'. Furthermore, the loading of each species on this factor is positively correlated with brain size, reflecting that larger-brained primate species, particularly the great apes, outperform smaller-brained species systematically across all cognitive domains (Deaner et al., 2006; Reader et al., 2011). However, such meta-analyses hinge on the quality of the published experimental studies. Unfortunately, the methods and results of experimental cognitive studies are often not reported in sufficient detail to be replicated accurately, and even small differences in methods within or between research facilities can lead to different outcomes. Moreover, although the term general intelligence is often used interchangeably for within-species (*g*) as well as between-species (*G*) comparisons, it is currently unknown whether between-species variation necessarily suggests within-species variation. Therefore, what is needed are valid and reliable comparisons of individuals within the same species.

To date, the most convincing findings regarding *g* still stem from research on human subjects (see Figure 1). However, the intriguing question is whether *g* is a uniquely human phenomenon or if there is continuity in the evolution of this domain-general cognitive meta-ability, i.e., do other primate species show similar consistent individual differences in cognitive performance?

Even though general intelligence at the species-level was originally thought to be an exclusively human trait, there is now accumulating evidence for *g* in several non-human animal species, suggesting evolutionary continuity, at least along the mammalian lineage and in some passerine birds (bower birds: Isden et al., 2013, Kaegy et al., 2011; and robins: Shaw et al., 2015). While most of this research has focused on rodents (mice and rats; e.g. Anderson, 1993; Galsworthy et al., 2002, 2005; Kolata et al. 2005, 2007; Light et al., 2008, 2010; Matzel et al., 2003, 2006) who can easily be kept and bred under captive conditions, evidence for *g* has more recently also been found in non-human primates, i.e., rhesus macaques (Herndon et al., 1997), cotton-top tamarins (Banerjee et al., 2009) and orang-utans (Damerius et al., 2017b), and the performance of chimpanzees across physical and social cognitive tasks also largely corresponds to this pattern (Hopkins et al., 2014; Woodley of Menie et al., 2015), with the exception of tasks containing a strong spatial component (Herrmann et al., 2010; see also van Schaik et al., 2012).

1.2.2 How is general intelligence measured in humans?

In humans, general intelligence is often measured with cognitive test batteries that consist of a number of cognitive subtests, i.e., nonverbal and verbal tasks that assess a wide range of different cognitive abilities. After being given verbal instructions, the test person has to solve cognitive tasks such as reasoning (e.g., identify which of several pictured abstract objects completes a logical series of similar objects), mental rotation (e.g., imagine which side of a 3-dimensional geometric figure would be visible after turning the figure in space to a certain number of degrees), memory (e.g., remembering the items of a 'shopping' list), or verbal fluency (e.g., generating as many words as possible that begin with a given letter). A widely used example of such a test battery is the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 2008).

Hundreds of large datasets on human intelligence, as measured by cognitive test batteries, support a hierarchical structure of cognitive abilities with general intelligence as an overarching global cognitive factor at the top (see Figure 1.1.).

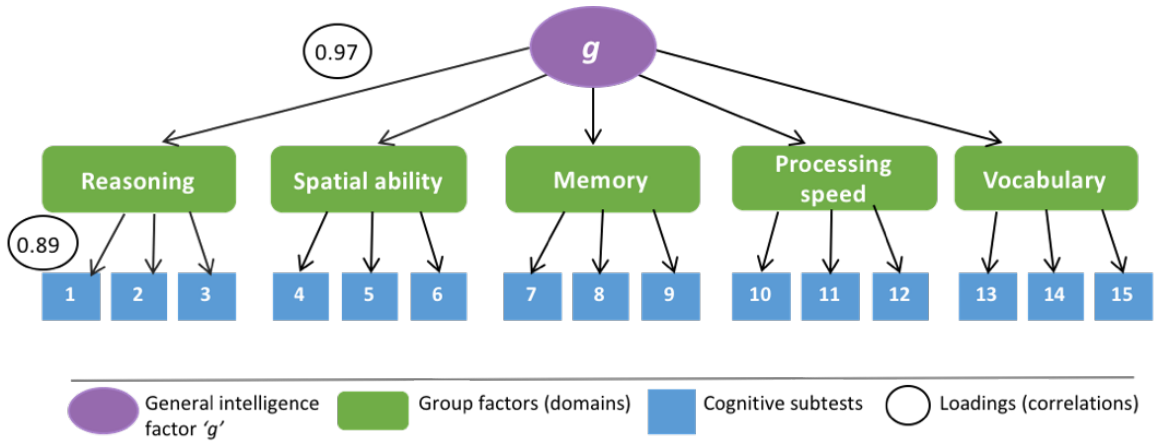


Figure 1.1 Measuring general intelligence 'g' in humans. Intelligence tests reveal a hierarchical structure of our cognitive abilities. The 16 different subtests measure a broad range of cognitive abilities. Some subtests load highly on group factors, that is cognitive domains such as spatial ability and memory. Loadings are indicated by numbers next to the arrows and represent the correlations between subtests and group factors (domains), e.g., three of the 16 subtests load equally high on the reasoning domain. Finally, all group factors are highly correlated with a single factor, the general intelligence factor *g*. (*Modified after Deary, Penke, & Johnson, 2010*).

Importantly, the cognitive test batteries available for human participants have regularly been evaluated regarding their validity (whether they truly measure the cognitive abilities they were designed to measure) and their re-test reliability (whether they produce repeatable results when administered twice). In addition, they have been standardised using large samples (usually thousands) of test persons from a variety of social and educational backgrounds and from different age groups (norms).

1.2.3 How is general intelligence measured in non-human animals?

During the last two decades, a large number of cognitive tasks and psychometric tests have been developed for non-human primates and they have often been widely used to compare a large number of different species. However, the majority of cognitive tests that have been developed and are currently available for non-human primates have not yet been evaluated regarding their measurement validity in a comparable way to cognitive tests available for humans.

Only in the last decade, researchers have started to test non-human primates with whole test batteries that are composed of several subtest that assess a wide range of cognitive abilities. The most prominent example is the Primate Cognition Test Battery (PCTB) (Herrmann et al., 2007) that consists of 16 subtests from two broad cognition domains, the physical and social one. The physical domain encompasses three scales that each

contain 1-4 different tasks: 'space' (spatial memory, object permanence, rotation, and transposition), 'quantities' (relative numbers and addition numbers), and 'causality' (noise, shape, tool use, and tool properties). The social domain entails three scales: 'social learning' (social learning), 'communication' (comprehension, pointing cups, attentional state) and 'Theory of mind' (gaze following and intentions). The PCTB was not specifically designed to assess general intelligence at the intra-specific level, but when the full test battery was used for this purpose, the findings were mostly negative (Herrmann et al., 2012, Amici et al., 2015) or inconclusive (Herrmann et al., 2010). However, when a subset of the tasks was used, the findings were mostly positive and in support of *g* (Hopkins et al., 2014; Woodley of Menie et al., 2014; see also Burkart, Schubiger, & van Schaik, 2017a).

Moreover, for non-human primates (and other animals), there are no norms to which the performance of a subject of a given species could be compared. Part of the problem is, that it is difficult to obtain large sample sizes, partly because researchers often do not have access to large enough numbers of non-human primate individuals, and partly because testing non-human animals is generally more demanding and time-consuming than testing human subjects.

The cognitive tests that are currently available for non-human primates compare best to criterion-referenced tests used for humans. Rather than being compared to other individuals, a subject's performance is compared to an objective criterion, usually how many per cent correct responses (e.g., choices between two or several options, one of which is classified as the correct choice) it has made out of a hundred per cent possible responses.

1.3 Measurement validity of cognitive test instruments

Despite the widespread lack of adequate validations, many tasks and the few existing cognitive test batteries are increasingly used for between-species comparisons. However, between-species comparisons rely on good experimental paradigms that provide truly comparable data within species and to date very little research has evaluated whether this is the case. Therefore, it is now time to call their internal or construct validity (Cronbach & Meehl, 1955) into question, that is whether cognitive tasks actually do measure what their human developers attempted to measure and to improve existing paradigms accordingly if this is not the case.

While construct validity is essential for within and across species comparisons of one or a small number of cognitive abilities, this may not be a comparable issue for within species comparisons of general intelligence that encompass tests from several domains. Nevertheless, although it may not be essential that each subtest measures the exact cognitive ability it was designed for, it is essential that it truly measures cognition rather than variation in other, non-cognitive, factors such as motivational ones.

1.3.1 The replication crisis in psychology

Although empirical researchers, such as psychologists and other scientists, can control many confounding factors in well-designed experimental studies, there is a considerable risk that a variety of biases potentially affect the outcomes of a study. In order to identify such potentially confounding factors, it is essential that empirical studies and their findings are replicated by researchers other than the original study authors. Recently, it became evident that replication of published findings it not always possible. When fellow researchers do attempt but fail to replicate research findings, even seemingly well-established ones, this can question not only the validity of the original findings but also whether the original authors used appropriate research practices, as illustrated by the 'replication crisis' in psychology (e.g., John et al., 2012; Makel et al., 2012; Pashler & Wagenmaekers, 2012).

There are many other factors that can affect the replicability of studies in psychology and in science in general. It is therefore essential for researchers to be aware of sources of potential biases and use objective research practices in order to reduce biases to a minimum and facilitate later replications of their results. Examples for such biases are that researchers tend to, often subconsciously (Newell & Shanks, 2014), design experiments and interpret the participants' behaviour in favour of their hypotheses (confirmation bias; Wason, 1960) or stop their data collection prematurely if a significant result is found, or conducting several small rather than one larger study.

One issue of psychological studies is that that sample sizes and thus statistical power tend to be relatively small. Sometimes this is even the case for psychometric studies as has been demonstrated by a meta-analysis on the positive correlation between how well pre-verbal infants habituated to stimuli during their first year of life and their IQ scores up to eight years of age (McCall & Carriger, 1993). Although all but one of the included studies

reported a significant positive correlation between habituation in early infancy and later child IQ, only one quarter of these studies had substantial statistical power and, contrarily to what would have been expected, the correlation was stronger in those studies with smaller sample sizes. One possible explanation was that several biases affected the results of those studies with smaller sample sizes, as Bakker et al. (2012) point out. For instance, since testing pre-verbal human infants is more demanding than older children and adults (e.g., clear verbal instructions are not possible), some participants might have been excluded in those studies based on arbitrarily chosen exclusion criteria and this in turn may have biased the results.

1.3.2 Replication crisis in comparative cognition research?

Contrarily to human psychology, the field of comparative cognition research has not (yet) been hit by the replication crisis to the same extent, but perhaps only because replication attempts are even more scarce than in other fields. In fact, it is very likely that comparative psychology and cognition research are even more prone to systematic biases that prevent replicability, because testing non-human primates and other animals is even more challenging than testing pre-linguistic human infants. While researchers can largely avoid many potential self-induced biases (such as confirmation biases) by using appropriate and objective experimental methods, the challenge of obtaining sufficiently large sample sizes remains because access to non-human individuals is often limited (Stevens, 2017). This is particularly the case for many primate species who are only housed in a few facilities around the world to which local research groups usually have priority access. Large sample sizes can often only be achieved by testing individuals from several different research facilities such as zoos or in larger sanctuaries where, sadly, non-human primates are more abundant owing to loss of their natural habitats (e.g., Damerius et al., 2017a, b; Herrmann et al., 2011). An alternative when large samples are not available is the replication of a study's findings in equally small but independent samples (as I have done in Study 2, *Chapter 3*).

Another factor particularly relevant when testing non-human individuals is the internal validity of cognitive studies (if they truly measure the dependant variables they intend to measure, rather than non-cognitive factors; Campbell, 1957). One important aspect, for instance, that may affect the validity and replicability of findings and which is often not considered appropriately in comparative cognition studies with non-human animals, is their

phenotypic plasticity (Voelkl & Würbel, 2016) at the species-and individual level. For instance, a subject's response to the experimental conditions not only depends on these conditions per se but also on the subject's internal state which is determined by its past and current exposure to environmental influences and its genetic predisposition to react to its environment in a particular way. As a consequence of this, different subjects may be at different points on the spectrum of possible states (termed the 'reaction norm', Woltereck, 1909) which affects how they respond during cognitive testing. However, while this inherent individual variation in behaviour may affect a subject's performance in a cognitive test, it is biologically meaningful and cannot simply be classified as measurement error. Since replications will be conducted under slightly different environmental conditions, particularly across different research facilities, it is therefore essential to be aware of and ideally assess this type of variation (as I have done in Study 1, *Chapter 2*).

The use of standardised experimental methods increases the internal validity of a cognitive study, including the construct validity of cognitive tests (how well they capture the construct they were designed to measure), and thus the reproducibility of a study. As I had mentioned earlier, a cognitive test's construct validity may not be a comparable issue for within species comparisons of general intelligence that encompass tests from several domains because this approach acknowledges that different cognitive abilities are not fully isolated from each other and some overlap between tests is to be expected (see also Huber, 2017; Ramus, 2017). However, although it may not be essential that each subtest measures the exact cognitive ability it was designed for, it is essential that it truly measures some aspect of cognition rather than non-cognitive factors such as emotional (as I have directly assessed in Study 1*, Chapter 2*) or motivational ones (as I have directly assessed in Study 1*, Chapter 2*, and indirectly in Study 2*, Chapter 3*).

While the internal or construct validity of cognitive tests is essential for within and across species comparisons of one or a small number of cognitive abilities, a study's external validity (how well the findings generalise to other circumstances such as environmental conditions or to other species, Campbell, 1957) is also important as it will influence large-scale comparative studies (across research facilities) and their replicability. I assessed the external validity of Studies 2 (*Chapter 3*) and 3 (*Chapter 4)* by exploring whether my findings generalised from my main study species to an evolutionarily closely related primate species.

1.4 Goal of this thesis: Assessing the influence of non-cognitive factors on cognitive performance

In a recent review paper, we provided an overview of the current state of comparative research regarding general intelligence and explored if the studies truly captured intelligence rather than statistical and methodological artefacts (Burkart, Schubiger, & van Schaik, 2017a; see *Appendix*). Furthermore, we suggested in what directions future research efforts could venture to extend our knowledge on general intelligence in nonhuman animals. We proposed a set of criteria that may be useful for future studies in order to (a) reliably determine if a given species possesses general intelligence, and (b) evaluate whether an experimentally and statistically determined *g* is a valid measure of intelligence in its broad definition (i.e., reasoning ability and behavioural flexibility) rather than just a methodological or statistical artefact as many researchers have claimed (e.g., Amici et al., 2012; Arden & Zietsch, 2017; Huber, 2017; Kovacs & Conway, 2017; Locurto, 2017; Shuker et al., 2017).

There is currently only a single cognitive test battery available that has been specifically designed to measure general intelligence in non-human primates and indeed found evidence for *g* in a New World monkey species, the cotton-top tamarin (*Saguinus oedipus*; Banerjee et al., 2009). However, like other test batteries, this battery has not yet been evaluated regarding whether it truly measures *g* rather than other factors that are not primarily cognitive in nature such as the subjects' motivation to attend and participate.

The aim of my PhD project was to take a first step into this direction by evaluating, in how far several non-cognitive factors affect cognitive performance in some of the subtests. The ultimate goal would be to investigate if a meaningful *g*-factor can be extracted in non-human primates and how we can best measure it. I tested common marmosets (*Callithrix jacchus*), close evolutionary relatives of tamarins, with an adaptation of five of these subtests (A-not-B, Detour-Reaching, Quantity Discrimination, Reversal Learning, and Memory; see Figure 1.2) in order to conduct a first evaluation and explore influential factors. All five tasks represented cognitive paradigms that have often been used to test non-human primates and involved a clear choice by the subject so that performance could be compared. The same applied to the additional memory task I had developed for Studies 2 and 3 in which both common marmosets and common squirrel monkeys (*Saimiri sciureus*) were tested with a subset of these cognitive tasks.

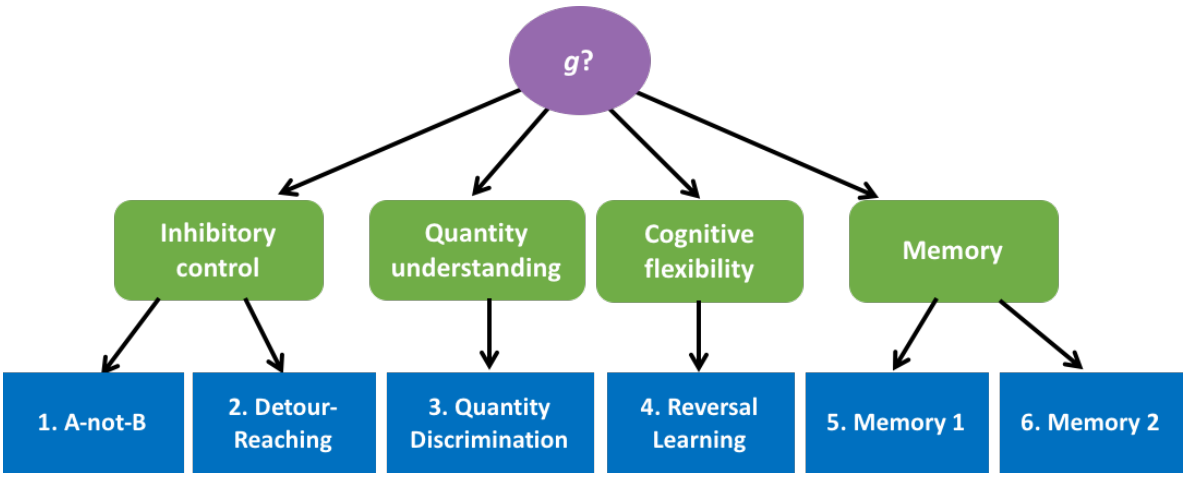


Figure 1.2 The cognitive test battery used in this thesis. The six subtests (blue rectangles) of our test battery with 1-2 tests per supposed cognitive domain (green boxes in the middle). Fifteen marmosets were tested with subtests 1-5, eight squirrel monkeys were tested with subtests 1, 2, 5, and 6, and 12 marmosets from an additional sample were tested with subtests 5 (two-choice version) and 6 (nine-choice version).

1.4.1 Thesis overview

Several criteria have been emphasised to warrant reliable results in comparative cognition studies (Burkart, Schubiger, & van Schaik, 2017a). In this thesis, I focused on several of the listed criteria (see Table 1.1). First, I ruled out that the subtests of the cognitive test battery measured factors other than purely cognitive ones such as state anxiety and motivation (*Chapters 2 and 3*). Second, while I prioritised testing the same individuals with all subtests and my sample size was therefore relatively small, the subtests consisted of diverse tasks tapping into several cognitive abilities (*Chapter 4*). Moreover, in contrast to other researchers (Herrmann et al., 2007; Banerjee, 2009), I did not categorise the cognitive tasks into a priori domains (*Chapter 4*). Third, I replicated the puzzling results of one task, the two-choice memory test, with two independent samples representing my study species and an evolutionarily closely related species (*Chapter 3*). Fourth, comparable to general intelligence studies with humans, I avoided tasks that are prone to secondary modularisation such as tool-use related ones, that would have measured cognitive abilities that appear to be evolutionary adapted (primary) modules but in fact need to be learned even by species with high general cognitive ability (secondary modules). Instead, I ensured that all subtests were as abstract as possible.

Table 1.1 A list of proposed criteria to avoid statistical and methodological artefacts to reliably identify a general intelligence factor *g* in nonhuman animals. The second and third column indicate to what extent these criteria have already been applied in rodents and primates, the last column indicates which criteria I applied in this thesis. (*Table after Burkart, Schubiger, & van Schaik, 2017*)

Criteria to avoid statistical & methodological artefacts	Applied in previous studies?		Applied in this thesis?
	Rodents	Primates	New World primates
1) Empirical control for confounds: • Anxiety • Motivation • Lower-level biological properties	Yes	No	Yes, in Chapters 2 & 3
2) Use of: • Diverse tasks • Statistical methods without an a priori categorisation of tasks into domains • Large samples	Mostly (least for task diversity)	Partly	Yes, in Chapter 4, apart from large samples
3) Replication of results in independent samples (when large samples are not available)	Yes	Partly	Partly, in Chapter 3
4) Avoidance of tasks prone to secondary modularisation	Yes	No	Yes, in whole thesis

1.4.2 Do individual differences in emotion affect motivation and cognitive performance? (*Chapter 2*)

The first study of my PhD project focused on the potential effects of individual differences in the subjects' emotional reactivity on their intrinsic motivation to participate and, more importantly, on their cognitive performance per se (see Figure 1.3).

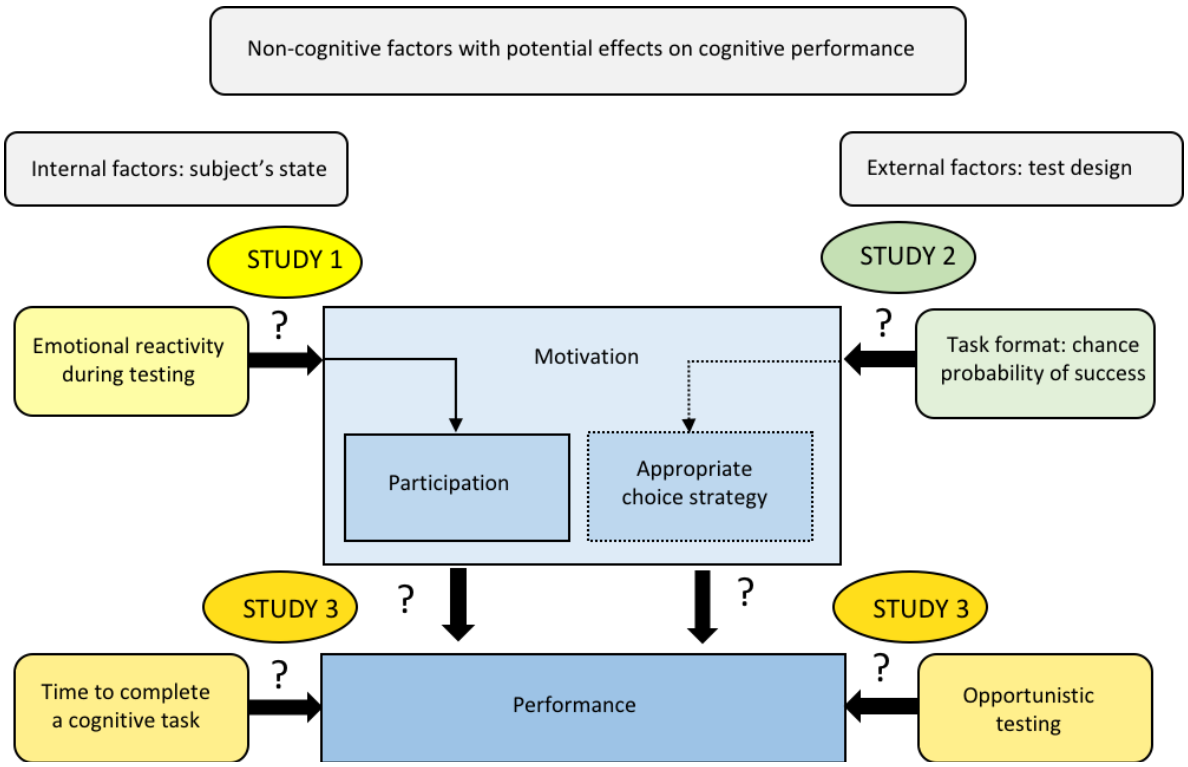


Figure 1.3 Overview of the aims of the thesis. Experimental assessment of three non-cognitive factors that could potentially affect the cognitive performance of the study subjects and lead to biased results. The dotted line and arrow in the motivation box indicate that the subjects' motivation to make correct choices was only indirectly assessed in Study 2 by varying the chance probability of success.

When testing nonhuman primates, including marmosets, some individuals appear at times emotionally aroused, inattentive and not highly motivated to participate in a cognitive test. In our facility, in such cases, a test session is terminated, the subject released back to its home enclosure, and the session continued on the following day. While most subjects re-engage in cognitive testing after applying this procedure, some still refuse to participate after such a break or continue to complete only a small number of trials per test session and testing day. Consequently, they eventually have to be excluded from further testing owing to ethical considerations (i.e., to avoid psychological distress) and time constraints (i.e., having to finish an experiment within a given time frame).

The aim of Study 1 (*Chapter 2*) was therefore to systematically quantify individual differences in the marmosets' emotional reactivity towards experimenters and the test situation, and to assess whether and how this affected the monkeys' participation and performance in a cognitive test (Figure 1.4). If individual differences in emotional reactivity lead to individual differences in the monkeys' cognitive performance, it would be essential not to exclude such individuals systematically from comparative studies. If, however, the

cognitive performance of highly emotionally reactive monkeys remained unaffected, including them in test batteries would be less essential. Moreover, if performance was not affected by the identity of the experimenter, the use of additional experimenters would be justified.

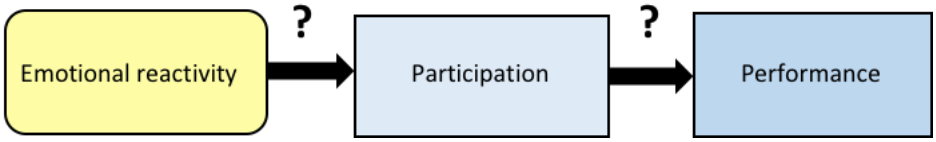


Figure 1.4 Study 1. Assessing the potential effect of individual differences in the subjects' emotional reactivity towards experimenter and test situation on their cognitive performance in an object permanence task.

As I will discuss in *Chapters 2 and 5*, I found considerable individual variation in the marmosets' emotional reactivity, and experimenters who were strangers elicited more arousal in the marmosets compared to highly familiar experimenters. Nevertheless, although highly emotionally aroused subjects tended to be less attentive and participated in fewer trials, their performance in an object permanence task was not affected.

1.4.3 How valid are standard memory tasks? (*Chapter 3*)

In my second study, I assessed and improved the construct validity (Cronbach & Meehl, 1955) of the memory test of the cognitive test battery. A cognitive test is regarded as construct-valid if it truly measures the cognitive ability (i.e. the cognitive construct) it was designed to measure, without additional factors, such as motivational aspects, significantly influencing an individual's test score.

The memory test in the classical two-choice task format was supposed to measure the marmosets' ability to remember, over increasingly longer time delays, the location of a food item visibly hidden in one of two containers. If this test validly measured long-term memory, I would have expected – in accordance with Ebbinghaus' forgetting curve - the marmosets' to have increasing difficulty to remember the reward's location over the course of the experimental sessions as a consequence of the increasing delays. However, in contrast to my expectation, I observed an initial peak, a medium low and a final peak in the subjects' supposed memory performance. One explanation for this pattern of results was that inhibition demands might have played a role in that it was most difficult for subjects to wait before being able to make a choice if the delay intervals had a medium duration. In any case,

the relatively low performance of the marmosets in general and the absence of a typical forgetting curve indicated that other factors than memory ability were determining the marmosets' performance in the task.

A more plausible explanation was that, overall, the marmosets performed poorly owing to the relatively low cost of using a random-choice strategy rather than remembering the reward's actual location. Even when choosing one of the two locations at random, a subject had a 50% chance per test trial that its choice was correct, allowing it to consume the food reward. If such a lack of motivation to choose correctly had caused the unexpected pattern of results, then lowering the possibility of random success by increasing the cost of a wrong choice (that would not allow the subject to consume the food item) should increase the subjects' motivation to choose correctly and lead to the expected forgetting curve. In order to further explore this possibility, I devised a new memory test version with nine rather than only two choice options. A new group of marmosets and squirrel monkeys was tested with both a two-choice and the nine-choice test version so that the performance in the two test versions could be directly compared in the two evolutionary closely related New World monkey species (Figure 1.5).

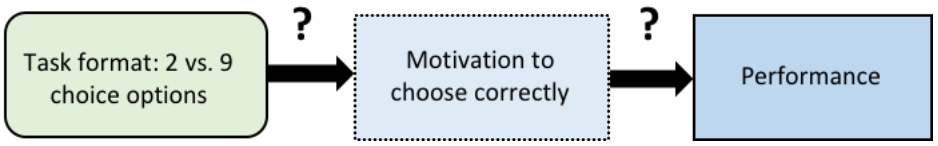


Figure 1.5 Study 2. Assessing the potential effect of the number of choice options (an aspect of the task format that affects the chance probability of success) on the subjects' motivation to use an appropriate choice strategy and their performance in a memory task.

As I will discuss in *Chapters 3 and 5*, both species performed better in the nine-choice than the two-choice memory test and, in accordance with my initial expectation, their performance in the nine-choice memory now decreased with increasing length of the time delay period, according to the expected forgetting curve.

1.4.4 Do individual differences in time to complete a cognitive test affect performance? (*Chapter 4*)

In my third study, I assessed whether individual differences in required testing time to complete a cognitive task lead to individual differences in cognitive performance. As I had already shown in my first experimental study (*Chapter 2*) conducted with marmosets, not all

subjects participate equally readily and consistently in cognitive tests. While all female subjects and some of the male subjects tended to be highly (food) motivated and maintained their motivation over the course of the numerous sessions of the test battery's subtests, other male subjects only participated in a few trials per session and required much more time to complete a test. The common practice in previous cognitive studies by other researchers has been to simply exclude those subjects with motivational issues from (further) testing and to select other subjects instead who participate more consistently in order to minimise testing time and maximise sample sizes. However, if these subjects are not motivated because the test is cognitively too demanding for them, excluding such subjects would heavily bias the results of the study.

To assess whether individual differences in required testing time affected the cognitive performance of marmosets and squirrel monkeys in the cognitive test battery (Figure 1.6), I used a full testing approach that gave the subjects sufficient time to complete the tasks at their own pace which reduced the number of dropouts. For each cognitive task, I compared the performance of subjects who completed the task within the expected number of testing days (and would have remained in the study sample had I used the classical opportunistic approach) to those subjects who needed longer (and would have been excluded in opportunistic testing). If the latter subjects differ from the former in terms of cognitive performance and ability, this would indicate that my full testing approach is not only advisable for general intelligence testing but for all types of cognitive testing. If, the two groups do not differ in terms of cognitive performance, this would indicate that opportunistic testing with numerous dropouts does not necessarily lead to a bias in the results.

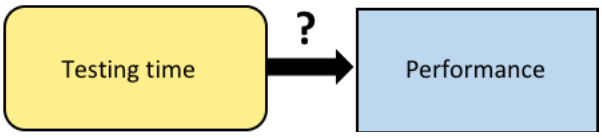


Figure 1.6 Study 3. Assessing the potential effect of individual differences in time subjects require to complete a cognitive task on their cognitive performance in a set of cognitive tasks.

As I will discuss in *Chapters 4 and 5*, I found that the two groups did not differ and that individual differences in testing time do not necessarily affect cognitive performance.

Chapter 2

High emotional reactivity towards an experimenter affects participation, but not performance, in cognitive tests with common marmosets (*Callithrix jacchus*)

Michèle N. Schubiger[1], Florian L. Wüstholz[2], André Wunder[3], & Judith M. Burkart[1]

[1]Anthropological Institute & Museum, University of Zurich, Switzerland

[2] Department of Philosophy, University of Fribourg, Switzerland

[3] Department of Theoretical Philosophy, University of Zurich, Switzerland

Published in: *Animal Cognition* (2015), *18*(3), 701-712.

2.1 Abstract

When testing primates with cognitive tasks, it is usually not considered that subjects differ markedly in terms of emotional reactivity towards the experimenter, which potentially affects a subject's cognitive performance. We addressed this issue in common marmosets (*Callithrix jacchus*), a monkey species in which males tend to show stronger emotional reactivity in testing situations, whereas females have been reported to outperform males in cognitive tasks. In a two-phase experiment, we first quantified the emotional reactivity of 14 subjects towards four different experimenters performing a standardized behavioural action sequence, and then assessed whether and how it affected the subjects' participation and performance in a subsequent object permanence task. A test session was terminated if a subject refused to make a choice in four consecutive trials. Highly emotionally aroused individuals, particularly males, were less likely to participate in the cognitive task and completed fewer trials. However, whenever they did participate and were attentive to the task, their performance was not affected. Our results suggest that differences in emotional reactivity towards an experimenter have no major impact on cognitive performance if strict criteria are applied on when to abandon a test session, and if performance is corrected for attention to the test procedure. Furthermore, they suggest that the reported sex differences in cognitive performance in marmosets may be owing to motivational and attentional factors, rather than a difference in cognitive ability per se.

2.2 Introduction

Whenever researchers conduct experiments with human or nonhuman primates, there is a risk that they will unintentionally influence the study's outcome in one or several ways. Consequently, the study subjects might change their behaviour in response to the experimenter rather than the experimental conditions. In order to avoid experimenter effects and to optimize comparability of results, different experimenters are usually trained to use identical standardized procedures. But can we ever be certain that different experimenters conducting the same test with the same subjects reliably obtain the same results in terms of the subjects' cognitive performance scores? One possibility is that different experimenters may elicit different emotional reactions in the subjects. In particular, some subjects may feel more at ease or more nervous with a specific experimenter than others. Such individual differences in emotional reactivity may influence a subject's motivation to participate in a cognitive test with this experimenter, and, importantly, this lower motivation in turn may affect their cognitive performance.

Possible effects of experimenters on subjects' performance scores in cognitive studies have been extensively studied in humans, and rodents, but only rarely been addressed systematically in non-human primates. The numerous reports of experimenter effects on the cognitive performance of human subjects include experimenters forming personal expectations on how different subjects might perform (e.g., Rosenthal, 1963) or differences in the way experimenters interact with their subjects such as being friendly or neutral (e.g., Siegwarth et al., 2012). In studies with rats, biases such as an experimenter's expectation regarding a subject's cognitive abilities or the degree of familiarity between the experimenter and the subject have been shown to potentially affect a subject's performance (see Schellinck et al., 2010). Systematic investigations of experimenter effects in non-human primates are rare and include two studies on capuchin monkeys, an observational field study on the putative effects of the presence of a human observer on the subjects' movement and activity patterns (Crofoot et al., 2010), and an experimental study that suggested experimenters may bias the results of cognitive tests by selectively choosing subjects with homogenous personality traits (Morton et al. 2013).

Crofoot et al. (2010) analysed the movement and activity patterns (recorded via telemetric collars) of seven habituated capuchin monkeys (*Sapajus apella*) in the presence and

absence of a human researcher. Overall, the capuchins did not change their movement and activity patterns in the presence of a human researcher. More subtle behavioural differences, however, could not be measured, such as the relative frequency of particular activities (e.g., vigilance, distress, or resting behaviour), and whether individuals differ systematically in such responses to the presence of a human researcher.

Morton et al. (2013) analysed whether the outcome of cognitive tests in captivity is affected by the so-called personality selection bias. In captive experiments, researchers can control for most confounding variables by careful study design. However, researchers may tend to selectively choose their study subjects based on particular personality traits, such as openness. This selection bias may ultimately result in a performance bias if such a personality trait is linked to cognitive performance. Morton and colleagues rated capuchin monkeys (*Sapajus apella*) on five personality dimensions and analysed their participation (how many sessions they completed) and performance (how fast they learned to choose correctly) in two training tasks. In the first task, food was positioned in front of one of two test compartments that the monkeys could freely access, and a correct response was to sit in the compartment that had the food in front of it. In the second task, a small opaque cup was placed in front of one compartment and a larger opaque cup in front of the other one, and a correct response was to sit in the compartment that had the larger cup in front of it. The authors found that individuals scoring higher on openness were more likely to participate and performed better in the first task, and assertiveness affected the monkeys' performance in both tasks, with less assertive subjects performing better than more assertive ones. Arguably, only the second task measured a truly cognitive component, and it is in this task that an individual's openness score no longer had an effect on its performance score. The negative effect of assertiveness may be best understood as individuals having motivational priorities other than food. A direct link between personality traits and cognitive ability per se could thus not be demonstrated unambiguously.

A study on rhesus macaques (*Macaca mulatta*, Toxopeus et al., 2005), however, showed that high levels on one dimension of trait anxiety, as assessed behaviourally in the group context in response to a loud noise, negatively impacted performance in a learning test. This result may suggest a link between one of three dimensions of trait anxiety and cognitive performance, but, as the authors stress, highly anxious subjects could simply be less attentive to the test than less anxious subjects, rather than having genuinely lower cognitive abilities.

This argument is particularly convincing since anxiety was also linked with low status: subordinate individuals may be less attentive to the task simply because they have to constantly monitor dominant conspecifics. In order to disentangle such confounding effects from genuine cognitive ability, it is necessary to control cognitive performance with regard to whether the subjects were attentive to the test procedure in the first place or not.

Importantly, excluding subjects from participation in cognitive tests is problematic, regardless of whether exclusions target individuals of a specific personality type, to the extent that they differ systematically in genuine cognitive performance. But even in the absence of such sampling biases, the risk of indirect and subtle experimenter influences on the subjects' performance remains. According to the Yerkes-Dodson law (Yerkes & Dodson, 1908), an individual performs best in a cognitive task if its emotional arousal is at an optimal level, which tends to be a medium degree of arousal. This optimal arousal level may vary both between and within individuals. While personality and trait anxiety may to some extent explain the origin of arousal-level differences between individuals, an individual's emotional arousal-levels also may vary from one cognitive task to the other, either as a consequence of task difficulty or a variety of other causes. Therefore, it is essential that the experimenter is aware of individual differences in the subjects' emotional reactivity and their possible consequences on its motivation, attention and cognitive performance. Ideally, a subject's behavioural signs of emotional arousal are evaluated prior to and during cognitive testing. The aim of our study was to systematically assess these issues in common marmoset monkeys (*Callithrix jacchus*).

Marmosets belong to the primates smallest in body size (Ford et al., 2009), which arguably makes them particularly likely to show high emotional reactivity toward human experimenters owing to the body size difference between experimenter and subject. Furthermore, reported and observed sex differences make them a particularly interesting study species to investigate the effect of emotional reactivity towards the experimenter on their participation and performance in cognitive tasks. In the wild, male marmosets are more vigilant than females (König et al., 1998), and in captivity, they seem to be more easily emotionally aroused during cognitive testing (personal observation). Emotionally aroused males often appear less willing to participate in a full test session or less attentive to the experimental stimuli when they do so. Intriguingly, male marmosets have also been reported to perform more poorly in cognitive tasks than females (Brown et al., 2010; Yamamoto et al., 2004). They thus represent an ideal test case to identify whether their poor performance

simply reflects emotional reactivity, a lack of motivation and attention or a true sex difference in cognitive ability. Yamamoto and colleagues (2004) tested female and male marmosets with a cognitive task that required the subjects to open an opaque plastic container by piercing its lid in order to obtain a raisin that they could see through a small hole in the lid. Male marmosets needed more time to approach and solve this task and used fewer strategies to obtain the reward. Brown et al. (2010) tested female and male marmosets' capacity to perceive biological motion by presenting them with point-light patterns of a walking hen on a computer screen. They found that female subjects paid more attention to the biological motion stimuli than males and spent more time inspecting them. The results of both above mentioned studies were interpreted as a sex differences in marmosets' cognitive abilities. However, an alternative explanation is that the poor performance of male marmosets found in these studies was an artefact of a lower motivation (e.g., to work for food) or lower attention resulting from higher emotional arousal in the test situation.

The aim of our study was to systematically assess individual differences in the subjects' spontaneous behavioural response to an experimental situation and their potential effect on their participation and performance in a cognitive task. In a two-phase experiment conducted in a single session, we quantified behavioural and bodily indicators of elevated emotional reactivity when marmosets were confronted with different experimenters and investigated whether variation in emotional reactivity towards an experimenter influenced the subjects' motivation to participate and their performance in a commonly used cognitive task.

In the first phase, an experimenter performed a standardized sequence of behavioural actions in front of the individual subject. This sequence consisted of the experimenter entering the room, performing a succession of six standardized actions, leaving and re-entering the room and finally offering a highly desirable reward to the subject. In the second phase, the same experimenter conducted an object permanence task with the subject, composed of a visible and an invisible displacement condition of six trials each. In order to maximize variation in emotional reactivity, we used different experimenters that were either familiar women (whom we expected to elicit low levels of emotional arousal) or unfamiliar men (whom we expected to elicit high levels of emotional arousal).

We addressed three main questions: (1) Do individual subjects react differently to different experimenters in terms of emotional arousal? (2) Are subjects with a stronger emotional response to an experimenter less willing to participate in a cognitive test?

(3) Do subjects with a stronger emotional response towards an experimenter perform worse in a cognitive task given they participate?

We expected to find substantial variation in the subjects' emotional reactivity and that male marmosets would react more strongly, particularly in response to unfamiliar experimenters. Furthermore, we anticipated that high emotional reactivity would lead to low participation whereas we had no prediction on whether high emotional reactivity would be linked to cognitive performance given we controlled for attention to the test stimuli and their manipulation. If males show higher emotional reactivity and lower participation as well as lower performance, even after controlling for attention, this would suggest genuine sex differences in cognitive ability. If, however, they showed higher reactivity, lower participation, but no difference in cognitive performance after controlling for attention, this would suggest that sex differences in cognitive performance are driven by motivational and attentional factors, rather than cognitive ability per se.

2.3 Materials and methods

2.3.1 Subjects

Study subjects were 14 adult captive common marmosets (*Callithrix jacchus*) of equal sex distribution, with a mean age of 7.11 years, living in family groups at the Primate Station of the Anthropological Institute of the University of Zurich. All marmosets were captive-born and mother-reared and participated in this study on four mornings after being fed their regular breakfast meal, a vitamin and calcium-enriched porridge. For a description of the study sample see Table 2.1.

Table 2.1 Description of study sample. Listed are the fourteen subjects' name, sex (7 females and 7 males) and age in years.

Subject	Sex	Age	Subject	Sex	Age
Jugo	Male	5.8	Thilo	Male	4.3
Juri	Male	11.0	Venezia	Female	6.5
Kaliper	Male	10.7	Venus	Female	7.7
Kapi	Male	9.8	Verona	Female	6.5
Tabor	Male	3.6	Vesta	Female	7.7
Tale	Female	4.3	Vito	Male	6.0
Tessy	Female	11.5	Vreni	Female	10.8

2.3.2 Experimenters

To maximize variation in the marmosets' emotional reactivity, we used four different human experimenters: two familiar women whom we expected to elicit low emotional arousal, and two unfamiliar men whom we expected to elicit high emotional arousal based on our experience with the study species (similar effects have recently been reported for rodents by Sorge et al, 2014). The marmosets had never seen the unfamiliar experimenters before this study but regularly participated in other studies with the familiar experimenters for at least a year. Note that the aim of this study was not to identify why marmosets react more strongly to some experimenters than others, i.e., to disentangle the effect of the experimenters' familiarity or gender on marmosets' emotional reactivity, which would have required counterbalancing the two factors. Rather, we were interested in the consequences of high emotional reactivity, regardless of its origin, for the marmosets' participation and performance in a subsequent cognitive task. Consequently, we chose experimenters whom we anticipated to produce the highest variation in the subjects' emotional reactivity. All four experimenters wore indiscernible black clothes and, most importantly, featureless black shoes as marmosets, an arboreal New World monkey species, tend to react strongly to obtrusive features at ground level (unpublished data).

2.3.3 Experimental set-up

All marmosets were tested individually in the same familiar experimental room and cage in which they had previously been tested in several socio-cognitive experiments (e.g., Burkart et al. 2007; Burkart et al. 2009). The experimental cage contained the actual test compartment (41 x 33 x 53 cm) and a directly connected second compartment of the same size. The two compartments were divided by a grid wall containing a rectangular opening (15 x 22.5 cm) in its bottom half through which the subject could move away from the experimenter and testing table and closer to its group members. We thereby ensured that very highly aroused subjects did not feel restricted in space. Each subject was tested individually in the same test compartment while its group members were waiting in an adjacent cage (100 x 78 x 122 cm). This waiting cage was connected to the experimental cage through a short gateway. Its lateral wall facing the experimental enclosure was covered with an opaque grey plastic board, its front with an opaque cream-colored fabric blind. This set-up prevented the waiting subjects from

seeing the experimenter until it was their turn to enter the test compartment while still allowing the current subject to hear and smell, but not see, its group members. Having such minimal 'contact' with group members is essential for marmosets' welfare as complete isolation from conspecifics imposes distress on this highly social species. Both cages were elevated 1.15 m from the ground as marmosets are arboreal and thus appear more comfortable in higher positions.

2.3.4 Apparatus

The basic experimental apparatus consisted of a testing table, a wooden board (40 x 40 cm) with two lateral plastic rails that was mounted on an adjustable tripod table so that the board's front was flush with the front of the experimental compartment. The wooden board of the testing table served as a base for the experimental equipment during the experimenter's behavioural action sequence (phase 1) and for the sliding platform and the cups during the cognitive task (phase 2). The equipment for the experimenter's behavioural action sequence consisted of two transparent glass containers (6 x 3 cm) with white plastic lids, one of which was half-filled with fine-grained pet sand, as well as of a digital timer and a dead cricket (*Acheta domesticus*; about 12 mm in size). The cognitive test apparatus itself consisted of a rectangular sliding platform (33.7 x 25 cm) made of cardboard that was covered with a washable wood-patterned plastic surface, and three cylindrical opaque black plastic cups (3.8 x 2.6 cm). The three aligned cups were placed at a distance of 3 cm from the platform's front edge, and the distance between the middle and each outer cup was 10 cm. All three cups were initially openly placed on the sliding board and kept in place with double-sided adhesive pads attached to the side that was lying on the board.

2.3.5 Experimental procedure

The experimental procedure consisted of two consecutive phases conducted in a single session with a maximal total duration of approximately 15 minutes that mainly depended on whether and how long an individual subject participated in the cognitive task. We used a within-subject design, where each of the four experimenters performed the two-phase experiment with all subjects. In phase 1, the experimenter (hereafter E) performed a series of six standardized behavioural actions in front of the subject that each lasted 30 seconds. In

phase 2, the E conducted a short object permanence task with the same subject. This resulted in 56 experimenter-subject dyads. Both phases, the standardized behavioural action sequence and the object permanence task, were identical for all subject-experimenter pairs. Experimenter order was counter-balanced across subjects to control for order effects. The two phases of the experiment are shown in Figure 2.1.

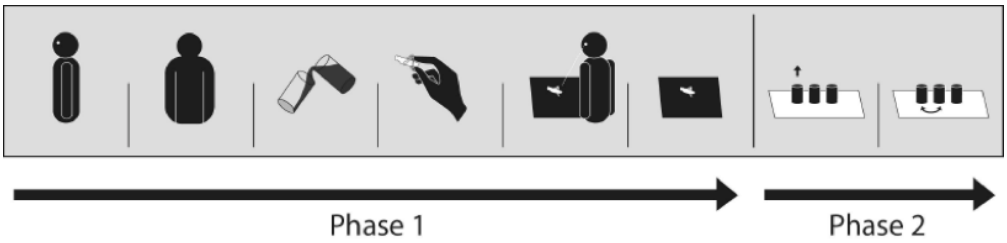


Figure 2.1 Two consecutive phases of the experiment. In phase 1, the experimenter (E) performed a sequence of six behavioural actions in front of the subject: (1) standing perpendicularly to it with head and eyes oriented away, (2) looking at the subject while oriented towards it and establishing eye contact, (3) manipulating an object (pouring sand between two transparent containers), (4) holding up a cricket (out of reach), (5) placing the cricket onto the test table/board (out of reach) and (6) leaving the test room (cricket still out of reach), re-entering it and giving the cricket to the subject. In phase 2, E conducted a cognitive task with the subject that consisted of two short object permanence tests of six trials each. In the first test, E placed the cricket under one of the three cups and closed all cups. In the second test, E again placed the cricket under one cup but then exchanged its location with an adjacent one. The subject made its choice by touching or lifting one of the cups.

Phase 1 – The standardised behavioural action sequence of the experimenter

The E entered the room, called the subject's name, walked towards the experimental cage on a marked line on the floor while looking down to this line and stopped 50 cm in front of the testing compartment, leaving the testing table between the E and the subject. The E then performed a standardized sequence of six behavioural actions, each lasting 30 seconds. A digital timer beeped to indicate when a 30 second interval had ended. For each action of the behavioural sequence, the E retrieved the equipment from a storage table at the back wall of the room and replaced it onto this storage table before starting the next action.

The experimenters' six sequential actions

(1) *Standing still*:

The E stood in front of the subject's test compartment. The E's body, head and eyes were facing 90 degrees away from the subject while the arms were in a neutrally stretched position on the sides of the E's body. After 30 seconds, the E turned towards the subject, and began with the second action.

(2) *Establishing eye contact:*

The E was now facing the subject and tried to establish eye contact for 30 seconds by following the subject with head and eye gaze.

(3) *Manipulation of an unfamiliar object*:

The E took the two small transparent containers (one was filled with sand, the other was empty) and placed them on the testing table. The E started the timer and removed the containers' lids, subsequently lifted up the containers, one in each hand, and visibly poured the sand from one container into the other and vice versa as soon as the first container was empty. The E repeated pouring the sand between the two containers, until 30 seconds had elapsed, closed the lids again and placed the containers back onto the storage table.

(4) *Holding food:*

After taking a cricket from the storage table, the E again stood in front of the testing table, as in the previous sequence, started the timer and held up the cricket visibly about 25 cm in front of E's chest without offering it to the subject. While doing so the E was facing the digital timer on the apparatus for 30 seconds.

(5) *Placing food out of reach/E present:*

The E laid down the previously held cricket onto the wooden platform at a distance of 17 cm (which is just out of the subject's reach) from the front of the test compartment and looked at the cricket for 30 seconds while standing still.

(6) *Placing food out of reach/E absent:*

The E left the room while the cricket remained placed on the wooden platform and out of the subject's reach. After the 30-second interval had elapsed, the E re-entered the room, picked up the out-of-reach cricket from the wooden board/test table and offered it to the subject.

Phase 2 - Object permanence test

Immediately after completing the behavioural action sequence, the E conducted a short object permanence task with the subject, consisting of two tests of six trials each, (1) a visible displacement test and (2) an invisible displacement test. The procedure corresponded principally to the one used by Mendes and Huber (2004) who found considerable individual variation in marmosets' performance. In the visible displacement test, the E placed the reward

twice under each of the three cups in a counterbalanced manner. In the invisible displacement test and in contrast to Mendes & Huber we did not use a transport cup and ensured that, at the time of the subject's choice, the reward appeared equally often in each location, twice on the left (A), twice in the middle (B), and twice on the right (C). To enhance comparability, all subjects first received six trials of the simpler test, the visible displacement. The invisible displacement test also consisted of six trials and followed immediately afterwards.

Task procedure

The E started the testing session by first placing the sliding platform with the three open black cups onto the test table. The E then called the subject's name, showed it a cricket, placed it in front of one of the open cups, and closed the cups from left to right.

Test 1: Visible displacement

In the visible displacement test, directly after baiting, the E slid the platform towards the test compartment so that the subject was able to make its choice by touching or lifting one of the cups.

Test 2: Invisible displacement

In the invisible displacement condition, the E baited one cup in the same way as in the visible displacement condition but performed a transposition immediately after closing all cups. The transposition consisted in the E exchanging the location of two cups, the baited and an adjacent empty one by simultaneously moving them on the board from one location to the other using the index and middle fingers of both hands. This resulted in the E's hands crossing, whereby the hand moving the baited container was always in the front. As in the visible displacement condition, the E then slid the platform towards the test compartment and allowed the subject to make its choice.

For both the visible and the invisible displacement test a choice was defined as the first cup the subject either lifted or touched. The procedure following the subject's choice was identical for both tests. If the subject correctly chose the baited cup, it obtained the cricket reward and the next trial started immediately. If the subject chose the wrong cup, the E drew back the platform, showed the subject the wrongly chosen container was empty and then opened the other two cups from left to right. The E then retrieved the cricket, placed it on the storage table, and continued with the next trial after a 15 sec time delay. If the subject refused to make a choice in a given trial within 30 seconds, the platform was also retrieved but no such

time out imposed. The next trial was directly started instead and the no-choice trial was repeated once all six trials of the test had been conducted, e.g., after the sixth trial of the visible displacement test.

Crucially, clear stop criteria were applied on when to terminate a test session. Testing was discontinued if a subject refused to choose any cup in four consecutive trials or if a subject did no longer make a choice in the last three trials of the invisible displacement test. This procedure ensured that subjects were allowed to leave the test situation if they were highly emotionally aroused, refused to make any choices or stopped choosing during the test session and thus appeared unmotivated. Therefore, while all subjects were presented with at least four of the six trials in the first test (visible displacement), not all subjects entered the second test (invisible displacement).

2.4 Data recording and analysis

2.4.1 Dependent measures during phase 1

In order to assess the extent of emotional reactivity in the subjects, we used several visible and audible behavioural indicators of emotional arousal usually shown by marmosets in stressful or challenging situations. We expected these indicators of emotional reactivity to be more pronounced with the unfamiliar than the familiar experimenters.

Behavioural indicators of arousal

(1) Tail-brush

In common marmosets, piloerection is most clearly visible in the brushing of the tail. A brushed tail indicates general vegetative arousal, but not necessarily of negative valence (Dettling et al. 2002).

(2) "Tsik" vocalisations

Tsik vocalisations are mobbing calls that marmosets typically emit in response to unfamiliar humans, stressful noises, other marmoset groups, and potential predators (Martins, Bezerra, and Souto 2008).

(3) "Egg" vocalizations

Egg vocalizations have been described as vigilance calls and are usually uttered if an unfamiliar human is approaching or if an individual marmoset finds itself in an exposed position

(Martins Bezerra & Souto, 2008).

(4) Self-scratching

Self-scratching has been described as a stress-associated behaviour for primates, including marmosets (Bassett et al, 2003).

(5) Escape

An escape involved the subject rattling and gnawing on the closed door (a grey plastic panel) on top of the experimental compartment in an attempt to leave the test situation.

(6) *Not taking offered food*

Not taking offered food was a subject's refusal to retrieve the cricket from the E's hand who offered it as a reward in the end of phase 1. From previous experiments, all subjects were used to retrieve rewards from an E's hand.

2.4.2 Data analysis

Both phases of the experiment, the behavioural action sequence of the E and the cognitive task, were video-recorded continuously with a digital camera. The subjects' spontaneous emotional response in the first phase was coded and analysed continuously from video recordings using the software package Interact from Mangold. Coding of phase 1 started as soon as the E started the digital timer and ended as soon as the subject had retrieved the cricket from the E's hand.

The definitions of the dependent variables are summarized in Table 2.2. The variables tail-brush (1) and escape (5) were measured as states. The beginning and end of a tail-brush state was determined by visible brushing or de-brushing of the tail, captured as a quick change (within a few seconds) in the degree of its piloerection. The duration of an escape state was delimited by the visible and audible start and end of the joggling and gnawing on the door at the top of the experimental compartment. The variables tsik calls (2), egg calls (3), self-scratching (4) and not taking offered food (6) were treated as events. Every individual egg and tsik vocalisation were coded as an event. Self-scratching was defined as an event of the subject using a hand or foot to scratch a part of its own body. Not taking the offered food was recorded as a single occurrence of the subject not retrieving the offered food from E's hand. The durations for the two state variables were measured in seconds and calculated as percentages of the total duration of phase 1. Events were calculated as relative frequencies by dividing the

absolute frequency of the behaviour by the total duration of phase 1.

Table 2.2 Definitions of the emotional reactivity measures. Four dependent variables (1,2,3, and 5) were included in the statistical analysis while the other two dependent variables (4 and 6 in italics) were excluded owing to floor effects.

	Dependent variable	Measurement level	Coded behavior
1	Tail-brush	State/duration	Visible change in piloerection of the tail
2	"Tsik" vocalisations	Event/frequency	Every individual call of a sequence was counted as a vocalisation
3	"Egg" vocalisations	Event/frequency	
4	*Self-scratching*	*Event/frequency*	*Using a hand or leg to scratch a part of its own body*
5	Escape	State/duration	Rattling and gnawing on the door at the top of the experimental compartment
6	*Not taking offered food*	*Single occurrence yes/no*	*No attempt to manually retrieve the offered cricket from the experimenter's hand*

2.4.3 Dependent measures during phase 2

Task participation was defined as the total number of trials the subjects completed in the visible and invisible displacement test.

The subjects' performance in the cognitive test was live-coded by the E who noted each of the subject's choices as correct, wrong or non-choice, and later verified these choices based on the video clips. Furthermore, we coded from the video clips for each trial in which a subject participated, whether the subject had actually seen the baiting and displacement or not, based on its head and body orientation during the manipulation of the experimenter. If the subject's head and eyes were oriented in a way that the E's hand and the containers were in its line of sight at the moment of the baiting and the transposition, a trial was coded as seen by the subject.

Task performance was calculated separately for each test and defined as the number of correct choices in all trials in which the subjects had paid attention to the presentation and therefore actually seen where the food was hidden or displaced.

2.4.4 Reliabilities and Statistical Data Analysis

In order to assess inter-rater reliability of behavioural coding for phase 1, and of task performance and attention in phase 2, 20% of the 56 video clips were coded by a second rater. Krippendorff's alpha coefficient was calculated for each dependent variable (Hayes & Krippendorff, 2007; Krippendorff, 2011). For the emotional reactivity variables from phase 1, alpha was generally high: egg calls $\alpha(12) = 0.99$, tail-brush $\alpha(12) = 0.91$, tsik calls $\alpha(12) = 0.97$, and escape $\alpha(12) = 0.99$. For phase 2, the cognitive task, all choices and non-choices were double-checked based on the video recordings. Correspondence between initially noted choices and double-checked choices was a 100%. Krippendorff's alpha for the subjects' attention in the first object permanence test was $\alpha(72) = 1.00$ for both the first and second object permanence task.

The dependent variables from phase 1 were first subjected to a principal component analysis (PCA without rotation) that established whether they represented the same or different dimensions of the subjects' emotional reactivity. However, self-scratching and not taking offered food were not used in this PCA owing to floor effects. Self-scratching was an extremely rare behaviour, and with the exception of two male subjects who refused to accept the food from one unfamiliar experimenter, all subjects accepted the offered reward from all four experimenters.

The principal components extracted from the PCA were used as explanatory factors in subsequent general linear mixed models (GLMMs) with task participation, task performance, and attention, as response variables, using the Standard Least Squares option and the restricted maximum likelihood method (REML). In all models, experimenter nested in experimenter group, and subject, were included as random effects. All statistical analyses were conducted using the software Jmp 10. We used a significance level of $\alpha = 0.05$ for all statistical tests, and all tests were two-tailed.

Is there individual variation in emotional reactivity towards different experimenters?

In order to answer the first question of whether the extent of the subjects' emotional reactivity differed when confronted with familiar versus unfamiliar experimenters, a separate GLMM was conducted for each of the two components resulting from the PCA (i.e., arousal and avoidance, see below).

Do individual differences in emotional reactivity affect the subjects' participation in the cognitive task?

To address the second question of whether elevated emotional reactivity (phase 1) influenced a subject's motivation to participate in the cognitive task (phase 2), we ran GLMMs for the whole object permanence task with the relative number of completed trials out of twelve as response variable. As fixed effects, we included arousal (as measured by factor 1 of the PCA), avoidance (as measured by factor 2), subject's sex, and their uni- and bivariate combinations with and without interactions. The best model was identified based on the lowest value of the Akaike information criterion, controlled for small sample sizes (AICc, Hurvich and Tsai, 1989). This model selection approach was chosen to account for the limited number of subjects.

Furthermore, we tested whether the extent of a subject's emotional reactivity affected how attentive it was during the baiting procedure in the first object permanence test and during the baiting and the transposition procedure of the cups in the second test. Therefore, we ran a second series of nine GLMMS with the same fixed and random effects as above, but the response variable was the number of trials a subject had attended to divided by all participated trials.

Do individual differences in emotional reactivity affect the subjects' performance in the cognitive task?

We analysed the third question of whether high emotional reactivity towards the experimenter and the test situation influenced the subjects' performance in the cognitive task by running several GLMMs separately for each of the two object permanence tests.

Analogous to question 2, we ran two series of nine different GLMMs all of which included the fixed effects subject's sex, arousal, avoidance and their interactions. Furthermore, we used two different response variables to elucidate how the subjects' attention to the test procedure influenced their performance, i.e., the total number of correct trials divided by the number of participated trials and the total number of correct trials a subject had attended to divided by all attended trials.

2.5 Results

2.5.1 Two independent dimensions of emotional reactivity

The PCA on the four emotional reactivity measures in phase 1 (Table 2.2) revealed two independent dimensions with Eigenvalues greater than 1 (2.167 and 1.186). Three dependent variables loaded highly on the first factor that explained 54.2% of the total variance in the dependent measures: egg calls (0.857), tail-brush (0.840) and tsik calls (0.837). Since all three variables represented audible and visible behavioural signs of arousal, the first factor was named "arousal". The fourth variable, escape, loaded very highly (0.954) on the second factor which explained 29.7% of the total variance and was named "avoidance" (see Table 2.3). To avoid collinearity issues, the two emotional reactivity factors, arousal (factor 1) and avoidance (factor 2), were used for all further statistical analyses.

Table 2.3 The two-factor structure of the subjects' emotional reactivity: arousal and avoidance behaviour. Loadings of the four dependent variables on the two extracted principal components (factors). Both stress-related vocalizations and piloerection of the tail (tail-brush) loaded highly on the first factor, arousal, which explained 54.2% of the total variation. The subjects' attempts to leave the situation loaded highly on the second factor, avoidance, which explained a further 29.7 % of the total variation. *High factor loadings are indicated in bold*

Dependent variable	F1 arousal	F2 avoidance
Egg calls	**0.857**	-0.287
Tail-brush	**0.840**	0.360
Tsik calls	**0.837**	-0.251
Escape attempts	0.161	**0.954**

2.5.2 Individual variation in emotional reactivity

A first GLMM with the response variable arousal and the fixed effects experimenter group and subject's sex, and their interaction, showed no significant main effect of experimenter group on subjects' arousal ($F(1, 2) = 12.88$, $p = .070$), although there was a trend for the subjects' arousal to be more pronounced when confronted with unfamiliar as opposed to familiar experimenters. There was, however, a significant effect of the subjects' sex $F(1, 12) = 7.28$, $p = .019$. A posthoc Student's t-test revealed that male marmosets showed significantly higher arousal than females $t(12) = 2.70$, $p = .019$ (Figure 2.2). Furthermore, the interaction between experimenter group and subjects' sex was highly significant

$F(1, 38) = 8.55$, $p < .006$. A Tuckey HSD posthoc test revealed that male marmosets showed significantly more emotional arousal when confronted with unfamiliar experimenters (Least Square *Mean* = 1.56, *SE* = 0.41) than with familiar ones (LS *Mean* = -0.22, *SE* = 0.41) and than female subjects with both unfamiliar (LS *Mean* = - 0.44, *SE* = 0.41) and familiar experimenters (LS *Mean* = - 0.90; *SE* = 0.41).

A second GLMM with avoidance as response variable revealed a strong trend in the same direction. Again, there was no significant effect of experimenter-group $F(1, 2) = 0.31$, $p = .634$ but a significant effect of the subjects' sex $F(1, 12) = 4.86$, $p = .048$. Male marmosets generally showed more avoidance behaviour than females $t(12) = 2.21$, $p = .048$ (Figure 2.2). However, although there was a trend for more pronounced escape behaviour of male marmosets in response to the unfamiliar experimenter group, the interaction between experimenter-group and subject's sex was not significant $F(1, 38) = 1.42$, $p = .241$.

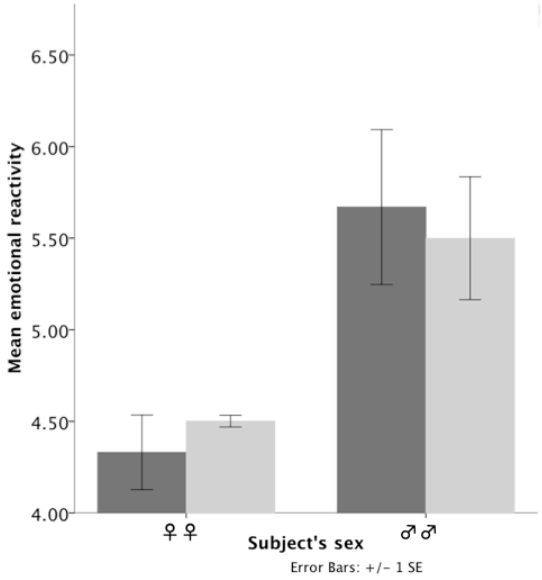


Figure 2.2 Effect of a subject's sex on its emotional reactivity. The dark grey bars represent the subjects' mean arousal (as measured by factor 1 of the PCA), the light grey bars their mean avoidance levels (as measured by factor 2). The two left bars represent female subjects (♀♀), the two right bars males (♂♂). *Note:* To better visualize the values that resulted from the PCA, they were transformed by adding a factor of 5 so that all values appeared greater than zero.

2.5.3 Emotional reactivity influences the subjects' participation

Participation in the object permanence task

The best model to explain the effect of the subjects' emotional reactivity on their overall participation in the object permanence task (both displacement tests) was the one that

included subject's sex as the only fixed effect (model 1, see Table S2.1 in the supplementary material). In this model a subject's sex had a highly significant effect on the number of completed trials $F(1, 12) = 41.16$, $p < .0001$. However, the model that additionally included arousal, showed nearly identical performance (model 4, ΔAICc = 1.840, see Table S2.1). In this model, both effects, a subject's sex ($F(1, 13.2)$, $p < .002$) and arousal ($F(1, 36.6)$ $p < .024$) significantly influenced participation in the test. A Student's t-test ($t(12) = 6.42$, $p < .0001$) revealed that male subjects completed significantly fewer trials (LS *Mean* = 0.46, *SE* = 0.09) than females (LS *Mean* = 0.99, *SE* = 0.09), who in fact all participated in all six trials of the visible displacement test and in most trials of the invisible displacement test (Figure 2.3a).

The relatively small difference between the two models (ΔAICc = 1.836) indicates that both variables, a subject's sex and its arousal, explain the individual variance in participation equally well. Their interaction was not significant (model 5, Table S2.1).

Attention in the object permanence task

The best model for the response variable attention, the number of seen trials out of all participated trials, was the one which only included the fixed effect arousal $F(1, 31.3) = 30.94$, $p < .0001$ (model 2, AICc = 40.780, see Table S2.2 in the supplementary material). Highly aroused subjects saw the test procedure in fewer participated trials than subjects with low arousal. Moreover, although there was no significant effect of avoidance on overall attention, avoidant subjects were less attentive in the first object permanence test ($F(1, 21.54) = 9.16$, $p = .006$ (Figure 2.4a).

2.5.4 Emotional reactivity does not influence cognitive test performance

Performance in test 1 visible displacement

When using the subjects' uncorrected performance as response variable, i.e., the number of correct trials of all participated trials in the first test, the best of the nine models based on the lowest Akaike information criterion was the one that included avoidance as the only fixed effect $F(1, 38.1) = 7.29$, $p = .010$ (model 3, ΔAICc = 18.035, see Table S2.3a in the supplementary material). Highly avoidant subjects made fewer correct choices than less avoidant ones (Figure 2.4b).

When using corrected performance as response variable, i.e., the number of correct

trials in which the subject saw the baiting of the cup, the best model was still the one with the fixed effect avoidance $F(1, 36.01) = 2.40$, $p = .130$ (model 3, AIC = 20.090, see Table S2.3b) but a subject's avoidance level no longer had a significant effect on its performance. Moreover, the model that only included subject's sex as fixed effect (model 1, ΔAICc = 0.315 see Table S2.3b) and the model including arousal only (model 2, ΔAICc = 1.876, see Table S2.3b) can be considered as equivalent based on their ΔAICc values lower than 2. None of these effects were significant.

Performance in test 2 invisible displacement

When we used the uncorrected performance as response variable, the best model included only subject's sex (model 1, AICc = -4.251, see Table S2.4a in the supplementary material). Male subjects tended to perform more poorly than females but this effect was not significant $F(1, 13.85) = 1.90$, $p = .189$.

When we used the corrected performance as response variable and thus controlled for attention, the best model was, as for test 1, the one including avoidance as single fixed effect $F(1, 18.95) = 1.05$, $p = .318$ (model 3, see Table S2.4b). However, the two models that only included sex (model 1, ΔAICc = 1.319) and arousal (model 2, ΔAICc = 0.368) resembled it closely (see Table S2.4). None of these effects were significant (Figure 2.3b).

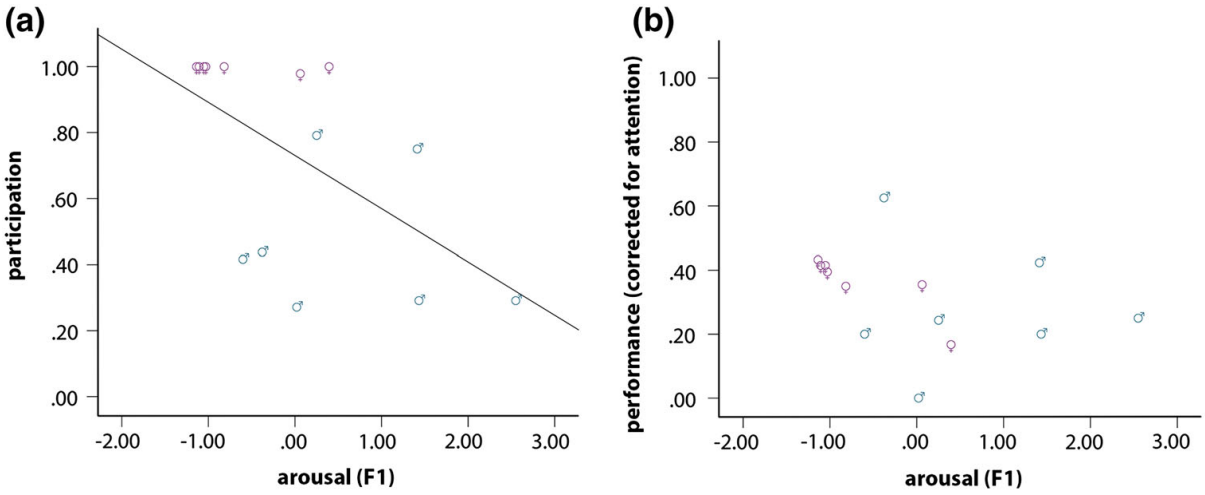


Figure 2.3 A subject's sex and emotional arousal affected its participation but not its performance in the object permanence task. Mean percentages of (a) participated trials of a subject in the object permanence task, and (b) correct trials in the second test, invisible displacement, corrected for attention. A subject's sex is indicated by its gender symbol (♀ = female, ♂ = male).

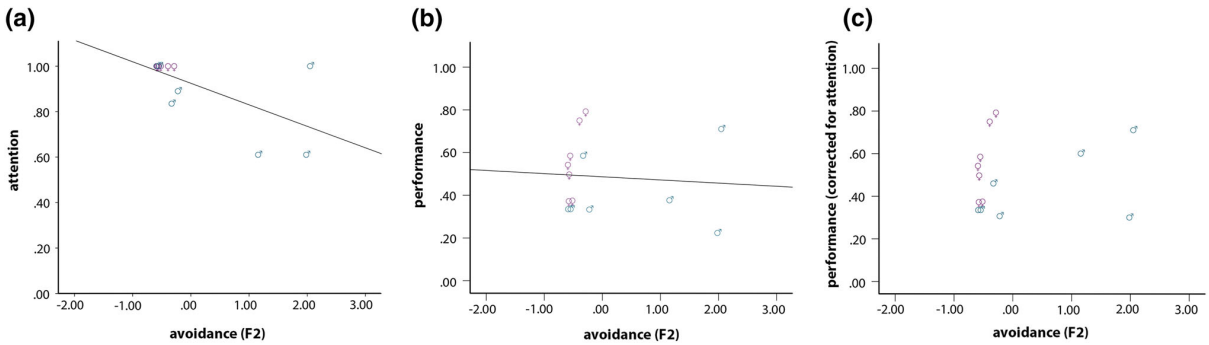


Figure 2.4 Emotional avoidance affected a subject's attention but not its performance in the first object permanence test, visible displacement. Shown is the effect of a subject's avoidance levels on (a) the mean percentage of trials it paid attention to, (b) its uncorrected performance: the mean percentage of correct of all participated trials, and (c) its corrected performance: the mean percentage of correct trials it paid attention to in the visible displacement test. A subject's sex is indicated by its gender symbol (♀ = female, ♂ = male).

2.6 Discussion

The aim of this study was to quantitatively assess whether and how a subject's emotional reactivity towards an experimenter and the test situation impacts participation and performance in a cognitive task.

In a two-phase experiment conducted in a single session, we quantified the extent of emotional reactivity of common marmosets towards four different experimenters. We then assessed whether and how emotional reactivity affected the subjects' participation and performance in a subsequent object permanence task. To maximize variation in the subjects' emotional reactivity, we confronted subjects with two groups of experimenters whom we expected to elicit high variation in emotional reactivity, two male experimenters, who were complete strangers to the subjects, versus two familiar female experimenters. In the first phase of the experiment, the experimenters performed an identical standardized sequence of behavioural actions in front of the subjects. We found considerable variability in the subjects' spontaneous emotional response as measured by four behavioural indicators that represented two independent dimensions of emotional reactivity, arousal and avoidance. The marmosets behaviourally expressed high arousal by showing piloerection of the tail and emitting two distress-related vocalizations, whereas avoidance was expressed by attempts to leave the experimental situation. Overall and in accordance with our expectation, male subjects were more emotionally aroused than females when confronted with the experimenters, in particular

with unfamiliar ones. They also tended to try to avoid the situation more often than female marmosets, independently from experimenter familiarity.

The extent of emotional arousal and avoidance during phase 1 had affected the male subjects' participation in the subsequent object permanence task in phase 2, whereas the participation of females was largely unaffected. Moreover, marmosets who had shown more attempts to leave the situation in phase 1 were less attentive to the test procedure when they eventually participated in the cognitive test.

In contrast to their participation, the subjects' cognitive performance in the object permanence task was not affected by emotional reactivity, given we controlled their performance for whether they had had paid attention to the experimental procedure or not. However, when not controlling for attention, avoidant individuals made more mistakes. Together, the results show that individuals who were avoidant in phase 1 were less likely to be attentive to the task prior to making their choice, and therefore made more mistakes. Thus, controlling for attention is vital in particular for individuals who show higher emotional reactivity. We conclude that in common marmosets, emotional reactivity towards an experimenter affects participation and attention but not performance in cognitive tasks. An interesting direction for future studies is to disentangle what factors are responsible for the higher emotional reactivity of some individuals. Besides a subject's sex, possible factors are its personality, its familiarity with the experimenter, the experimenter's gender, similar personalities of both the subject and the experimenter, or a combination of these factors.

In our study, male marmosets showed a stronger emotional response, were less motivated to participate, completed fewer trials, and were less likely to attend to the presentation and experimental manipulation of the cups and the reward. However, in the trials in which they did participate and paid attention to the presentation and manipulation of these stimuli, their cognitive performance did not differ from their female conspecifics' performance. This result contrasts with the idea that female common marmosets possess better cognitive abilities than males (Yamamoto et al, 2004; Brown et al, 2010). Our results suggest that these reported sex differences in cognitive performance in marmosets are best understood in terms of motivational and attentional sex differences, such as a higher interest of female marmosets to work for food and of male marmosets to remain vigilant to the environment (see also König et al. 1998), rather than reflecting a true sex difference in their cognitive abilities.

Our results have two implications for conducting experimental tests with marmosets, and arguably nonhuman primates in general. First, comparable results may be achieved with different experimenters, even if subjects vary individually in how they behaviourally respond to these different experimenters. It is crucial, however, to control for whether the subjects are attentive to the presentation and manipulation of the experimental stimuli, because particularly highly reactive individuals may participate in the task without paying attention to it, perhaps in order to quickly finish the experiment. Ideally, the experimenter should obtain the subject's full attention before starting a test trial and closely attend to whether the subject stays attentive during the whole test procedure. However, monitoring a subject's attention during testing can be difficult, because the experimenter has to avoid to inadvertently give visual cues to the subject and cannot look at it during the actual test trial. Therefore, it is vital to determine from video recordings for every trial if the subject had been fully attentive or not, and to correct its performance accordingly.

The second implication is that excluding highly reactive animals from testing does not necessarily bias the results. In fact, the alternative explanation of the trait-anxiety effect in macaques (Toxopeus et al., 2005), i.e., that the adverse effect of trait-anxiety on cognitive performance results from attentional biases rather than cognitive ability per se, is supported by the present set of findings. Likewise, our results are consistent with the ones obtained by Morton et al. (2013) who basically showed a participation effect rather than a truly cognitive one.

Based on our findings, we suggest that differences in emotional reactivity do not necessarily have an impact on cognitive performance. Thus, it is justifiable both to use different experimenters as well as to exclude very highly aroused and unmotivated subjects from testing. However, observable behavioural indicators of elevated emotional arousal should sensitise the experimenter to evaluate an individual subject's attentional state prior to and during cognitive testing.

2.7 Acknowledgements

We thank Prof. Carel van Schaik for discussion, Dr. Karin Isler and Christa Finkenwirth for statistical advice, Caroline Jordan for the graphic implementation of Figure 1, and our animal keepers, Heinz Galli and Thomas Bischof for help with the preparation of the experimental room prior to this study. We also thank two anonymous reviewers for their helpful comments on the initial manuscript.

The study was funded by the Swiss National Science Foundation (project number: 310030_130383). It was performed under guidelines established by the National Veterinary Office of Switzerland and licensed to be performed by the Veterinary Office of the Canton of Zürich. The authors declare they have no conflict of interest.

2.8 Supplementary material

Tables S2.1 – S2.4 provide an overview of the results of the GLMMs computed in JMP 10 to determine the effects of a subject's emotional reactivity and sex on its participation, attention, and performance in the object permanence.

Shown are the models 1 to 9, the p-values for each included fixed, the corrected Akaike information criterion (AICc), and its difference to the best model (ΔAIC).

Table S2.1 Effect of a subject's emotional reactivity on its overall participation in the object permanence task. The response variable was the number of participated trials in both displacement tests. Model 1 which only includes the subjects' sex provides the best fit based on the lowest AICc. However, model 4 which also includes the subjects' arousal can be considered as equivalent based on the low ΔAIC < 2.

Model	Sex	Arousal	Avoid	Sex*Arousal	Sex*Avoid	Arousal*Avoid	AIC	ΔAIC
1	< .0001***						47.393	0
2		.0007**					52.523	5.130
3			.038*				61.085	13.693
4	.0018**	.0238*					49.229	1.840
5	.0014**	.315		.213			54.534	7.142
6	.0003***		.736				54.207	6.815
7	.019*		.679		.741		58.580	11.187
8		.0002**	.010**				50.265	2.872
9		.018*	.014*			.770	56.890	9.500

Table S2.2 Effect of a subject's emotional reactivity on its attention in the participated trials of the object permanence task. The response variable was the number of participated trials that were actually seen by the subject in both tests. The best model based on the AICc is model 2 that only includes the subjects' arousal.

Model	Sex	Arousal	Avoid	Sex*Arousal	Sex*Avoid	Arousal*Avoid	AIC	ΔAIC
1	<.0001***						46.725	5.945
2		.0001**					40.780	0
3			.651				62.401	21.621
4	.063	.001**					43.426	2.646
5	.040*	.056		.198			48.621	7.841
6	<.0001***		.253				52.250	11.469
7	.026*		.859		.610		56.511	15.731
8		<.0001***	.590				47.266	6.486
9		.024*	.326			.353	53.138	12.357

Table S2.3 Effect of a subject's emotional reactivity on its performance in the visible displacement test. The response variable was the number of correct out of all participated trials of a subject (a) without and (b) with correction for attention.

(a) Performance in visible displacement, not corrected for attention

The best fitting model is Model 3. Highly avoidant subjects performed worse in the test than subjects who showed less avoidant behaviour. Note that no p-values could be computed for model 4.

Model	Sex	Arousal	Avoid	Sex*Arousal	Sex*Avoid	Arousal*Avoid	AIC	ΔAIC
1	.176						22.894	4.345
2		.911					24.603	6.568
3			.010*				18.035	0
4	-	-					29.138	11.103
5	.161	.627		.752			36.162	18.128
6	.962		.018*				24.622	6.356
7	.864		.440		.854		29.622	11.588
8		.585	.010*				24.701	6.667
9		.166	.002*			.077	27.603	9.566

(b) Performance in visible displacement, corrected for attention

Model 3 is still the best model but the effect of a subject's avoidance level on its performance in the visible displacement test was no longer significant after excluding trials in which the subjects had not been attentive to the test procedure. Model 1 and 2 can be considered as equivalent based on their ΔAICc values < 2.

Model	Sex	Arousal	Avoid	Sex*Arousal	Sex*Avoid	Arousal*Avoid	AIC	ΔAIC
1	.247						20.404	0.315
2		.721					21.966	1.876
3			.130				20.090	0
4	.191	.454					26.813	6.724
5	.193	.428		.757			33.912	13.832
6	.646		.248				26.330	6.240
7	.600		.795		.839		31.487	11.397
8		.559	.112				26.690	6.003
9		.297	.062			.284	31.457	11.367

Table S2.4 Effect of a subject's emotional reactivity on its performance in the invisible displacement test. The response variable was the number of correct out of all participated trials of a subject (a) without and (b) with correction for attention.

(a) Performance in the invisible displacement, not corrected for attention

The best fitting model is Model 1. Males showed a non-significant trend to perform worse in the invisible displacement test given they participated. But model 3 can be considered as equivalent based on Δ<2. Note that no p-values could be computed for model 5.

Model	Sex	Arousal	Avoid	Sex*Arousal	Sex*Avoid	Arousal*Avoid	AIC	ΔAIC
1	.189						4.251	0
2		.842					2.075	2.175
3			.674				2.711	1.540
4	.176	.547					3.344	7.595
5	-	-		-			9.250	13.501
6	.225		.796				2.909	7.160
7	.780		.028*		.011*		1.034	5.285
8		.912	.695				5.125	9.375
9		.571	.193			.210	9.501	13.752

(b) Performance in the invisible displacement, corrected for attention

The response variable was the number of correct out of all participated trials that were seen by the subject. Model 3 which included only the fixed effect avoidance was the best model. But model 1 and 2 can be considered as equivalent based on Δ<2.

Model	Sex	Arousal	Avoid	Sex*Arousal	Sex*Avoid	Arousal*Avoid	AIC	ΔAIC
1	.694						15.676	1.319
2		.203					14.726	0.368
3			.318				14.357	0
4	.835	.269					22.335	7.977
5	.908	.146		.328			29.410	15.053
6	.825		.365				21.481	7.123
7	.269		.060		.968		23.226	8.869
8		.458	.479				21.500	7.143
9		.427	.874			.008	27.386	13.209

Chapter 3

How task format affects cognitive performance: a memory test with two species of New World monkeys

Michèle N. Schubiger[1,2], Alexandra Kissling[1], & Judith M. Burkart[1]

[1]Department of Anthropology, University of Zurich, Zurich, Switzerland

[2]Division of Psychology, Abertay University, Dundee, United Kingdom

Published in: *Animal Behaviour* (2016), *121*, 33-39.

3.1 Abstract

In cognitive tests, animals are often given a choice between two options and obtain a reward if they choose correctly. We investigated whether task format affects subjects' performance in a physical cognition test. In experiment 1, a two-choice memory test, 15 marmosets*, Callithrix jacchus*, had to remember the location of a food reward over time delays of increasing duration. We predicted that their performance would decline with increasing delay, but this was not found. One possible explanation was that the subjects were not sufficiently motivated to choose correctly when presented with only two options because in each trial they had a 50% chance of being rewarded. In experiment 2, we explored this possibility by testing eight naïve marmosets and seven squirrel monkeys, *Saimiri sciureus*, with both the traditional two-choice and a new nine-choice version of the memory test that increased the cost of a wrong choice. We found that task format affected the monkeys' performance. When choosing between nine options, both species performed better and their performance declined as delays became longer. Our results suggest that the two-choice format compromises the assessment of physical cognition, at least in memory tests with these New World monkeys, whereas providing more options, which decreases the probability of obtaining a reward when making a random guess, improves both performance and measurement validity of memory. Our findings suggest that two-choice tasks should be used with caution in comparisons within and across species because they are prone to motivational biases.

3.2 Introduction

When the cognitive abilities of animals are assessed with cognitive tests, subjects are often presented with two options to choose from and rewarded with a food item if they choose the correct option. This two-choice task format has been used to test, in a range of animal species, a variety of cognitive abilities such as memory (e.g., delayed response tasks in bees, *Apis mellifera*; pigeons, *Columba livia*; several rat strains; many other species, including primates; reviewed in Lind, Enquist, & Ghirlanda, 2015), understanding intentional deception (chimpanzees, *Pan troglodytes*, Woodruff & Premack, 1979; dogs, *Canis familaris*, Petter, Musolino, Roberts & Cole, 2009) or inferential reasoning (dogs, Erdöhegyi, Topál, Virányi, & Miklósi, 2007; carrion crows, *Corvus corone corone*, Mikolatsch, Kotrschal, & Schloegel, 2012; chimpanzees, bonobos, *Pan paniscus*, orang-utans, *Pongo pygmaeus*, gorillas, *Gorilla gorilla*, Call, 2006). One test that has extensively used the two-choice format in particular with a wide range of animal species is the object choice task. This task tests for sociocognitive abilities by assessing a subject's ability to use an experimenter's gestural cues (e.g., gaze, point, touch) in order to locate a reward that is hidden under one of usually two containers. The tested species include primates (all four great apes and some Old and New World monkeys), domesticated mammals (dogs; foxes, *Vulpes vulpes*; cats, *Felis catus*; horses, *Equus caballus*; goats, *Capra hircus*) and undomesticated terrestrial (wolves, *Canis lupus*; bats, *Pteropus spp.*) and marine mammals (dolphins, *Tursiops truncatus*; seals, *Halichoerus grypus* and *Arctocephalus pusillus*; sea lions, *Otaria byronia*), corvids (jackdaws, *Corvus monedula*, nutcrackers, *Nucifraga columbiana*) and parrots (African grey parrot, *Psittacus erithacus*); see Mulcahy and Hedge (2012) for a review.

Although the two-choice task format is widely used in comparative psychology, there is recent evidence that in some circumstances the task may not be a suitable method for assessing cognitive abilities. Burkart and Heschl (2006), for instance, found that common marmosets, *Callithrix jacchus*, a New World monkey species, chose at random when presented with only two containers in an object choice task, but they were able to use the experimenter's cues much more reliably and made more correct choices when presented with nine instead of only two containers to choose from. A likely explanation is that lowering the probability of obtaining a reward by random choice helped the marmosets to overcome an inherent social bias that makes nonhuman primates reluctant to follow communicative cues to food rewards.

In physical cognition tasks, such social biases should not influence a subject's performance, because these tasks usually do not involve any social interaction between subject and experimenter. Memory tests, such as delayed response tasks (e.g., Kendrick, Rilling, & Denny, 1986; Lind et al., 2015; Rodriguez & Paule, 2009) for instance, often require the subjects to first observe and later remember in which of two locations a reward has been hidden without obtaining any communicative cues. Consequently, if social biases alone were responsible for the effect of task format on the marmosets' performance in the object choice task, lowering the chance probability of success should not affect their performance in such non-social cognition tasks. Nevertheless, the subjects may prefer to choose in a random manner for other reasons, for instance to avoid the effort of memorizing. To date, it is not known whether, or to what extent, task format and chance probabilities also affect performance in physical cognition tests. But if they do so in a similar way, as demonstrated for social tests, this has far-reaching consequences for the validity of species comparisons that are often based on tasks that differ in format.

In the present study, we tested New World monkeys with a physical cognition test that assesses their memory ability and investigated whether an alternative task format with nine choices would also be more suitable than the traditional two-choice task format. In experiment 1, we tested common marmosets with a traditional two-choice memory test, i.e. the memory subtest (hidden reward retrieval) of a cognitive test battery designed to assess general intelligence in nonhuman primates (Banerjee et al., 2009). In this traditional delayed response memory test, the subjects had to remember the location of a food reward over various time delays. After watching how a food reward was hidden in one of two locations, the subject could no longer see the reward and had to wait until the delay interval had expired before it could choose one of the two locations. New World monkeys, particularly smaller species such as marmosets (Miles, 1956; Miles, 1957a) and squirrel monkeys, *Saimiri sciureus* (French, 1959; Miles, 1957b), have been shown to perform worse on such delayed response tasks than Old World monkeys (mainly rhesus macaques, *Macaca mulatta*) and apes (e.g., Fischer & Kitchener, 1965; Harlow, 1932; Miles & Meyer, 1956; reviewed in: Tomasello & Call, 1997). Even though the methodological details are not always comparable, New World monkeys have also been shown to perform as well as (capuchins, *Cebus apella*) or better than (spider monkeys, *Ateles geoffroyi*) Old World monkeys (long-tailed macaques, *Macaca fascicularis*), and even as well as great apes (Amici, Aureli, & Call, 2010). Moreover, even

smaller monkeys usually still perform well above chance, at least with short delays (comparison of apes and monkeys, Fischer & Kitchener, 1965). We therefore expected the marmosets to pass the traditional memory test in experiment 1. Furthermore, in humans, ability to remember a specific memory content declines exponentially the more time has elapsed since its acquisition, a phenomenon known as the forgetting curve (Ebbinghaus, 1885, 1913; hereafter Ebbinghaus effect). In experiment 1, we therefore expected that the marmosets' performance would similarly decline with increasing duration of the time delay if this test accurately measured memory performance. Since the marmosets performed relatively poorly in experiment 1 and did not show an Ebbinghaus effect, we conducted experiment 2, which was designed to assess the effect of reducing the chance of obtaining a reward when choosing at random. We tested a new sample of marmosets and squirrel monkeys and compared their performance in a traditional two-choice versus our newly developed nine-choice version of the memory test.

3.3 Experiment 1: Traditional two-choice memory test

3.3.1 Methods

Subjects

Fifteen common marmosets, eight males and seven females, participated in this study. All subjects were housed in social groups consisting of two to six individuals at the Primate Station of the Department of Anthropology of the University of Zurich, Switzerland. Their indoor enclosures had both daylight and artificial light and were composed of one to three components (depending on group size) measuring 1 x 2 m and 2 m high, each of which was equipped with several climbing structures such as natural branches, a sleeping box, an infrared lamp and a mulch floor. Whenever the weather conditions allowed it, each group had free access to an outdoor enclosure. The marmosets were fed a vitamin and calcium-enriched porridge in the morning, fresh fruit and vegetables at lunchtime, and gum and mealworms in the late afternoon. In addition, they received a daily protein snack in the afternoon such as pieces of cooked egg. Water was available ad libitum from water dispensers. All subjects were tested between their regular feedings and never food deprived during the study. They could

enter and leave the test enclosure through semi-transparent plastic tubes that were connected to their home enclosures and were not handled at any time.

Materials and Set-up

Each subject was tested individually in the same compartment (41 x 53 cm and 33 cm high) of a larger test enclosure, with its group members present in an adjacent enclosure (100 x 122 cm and 78 cm high) so that the subject could hear and smell but not see them during testing. The test compartment had a transparent Plexiglas window front containing two rectangular openings (4 x 2.5 cm). The test apparatus consisted of two white opaque cylinder-shaped plastic containers (3.0 cm in height and 5.3 cm in diameter) that were attached to a wooden board (33 x 33 cm) placed 2 cm from its front, and was placed on the wooden test table (40 x 40 cm) that was level with the test compartment's floor. The test apparatus could be slid in and out of the subject's reach. The two containers were filled with dark-brown bark mulch that corresponded to the flooring substrate in the marmosets' home enclosures. A small piece of a yellow locust, *Schistocerca gregaria*, served as a reward in each trial. At the beginning of each trial, the test apparatus was placed just out of the subject's reach and the two containers were each covered with a rectangular piece of mulch approximately the same size as the container.

Procedure

The experimenter stood behind the test apparatus, called the subject's name, said 'look' while showing it the reward and started a trial as soon as the subject was attentive. She removed the cover of one of the two containers, placed the food reward in the container and again covered it with the piece of mulch so that the reward was no longer visible and both containers, the baited and the empty one, remained covered. After the delay interval had expired, she slid the board with the containers towards the test compartment's window. The subject could then make a choice by reaching through one of two rectangular openings in the window and removing the cover with its hand(s). There were six conditions with increasing time delays of 5, 10, 15, 20, 25 and 30 s. Each test session consisted of 10 trials of one delay condition, if possible, conducted on the same day, which resulted in a total of 60 trials per subject. The reward's location was counterbalanced in a pseudo-randomized manner so that a piece of locust was hidden five times in the left and five times in the right-hand container but

never in the same container in more than two consecutive trials. Prior to entering the actual test sessions, each subject went through a pretest phase in which the experimenter followed the same procedure but did not impose a time delay. After the subject reached criterion (≥ 80% correct choices within a single pretest session of 10 trials), it entered the test phase. At the beginning of each test day, the subject received one warm-up trial, again without a time delay. Once a subject had finished the six test sessions, it was retested with one full session without a delay. If the subjects had understood the task, we expected their performance in this retest session to be higher than or at least as high as in the test sessions because the retest involved no memory demand. We used two predefined criteria for stopping a test session: (1) the subject did not make a choice in three consecutive trials, and (2) the subject was no longer attentive (not looking at the test apparatus but vigilant towards its surroundings instead) to the task, and/or emotionally aroused (producing vocalizations of discomfort and showing piloerection of the tail; for definitions see Schubiger, Wüstholz, Wunder, & Burkart, 2015), and indicated it wanted to leave the test compartment (climbing to and rattling on the door on top of the test compartment). If the subject met at least one of these criteria, it was allowed to go back to its home enclosure and the session was continued the following day.

3.3.2 Data scoring and analysis

Of the 12 subjects who completed all test sessions, one male subject (Jugo) only completed five trials of the retest and a second male subject (Vito) did not participate in the retest. Three subjects, two males (Kapi and Kantor) and one female (Kitty), did not complete the whole test phase, which resulted in a final total trial number of 756.

All trials were video recorded. The experimenter coded the subjects' choices live using check sheets and checked all trials a second time using the video clips. Five trials (0.7%) had to be excluded from the analysis owing to ambiguous behaviour of the subject or experimenter error. A second rater coded 20% of the trials from videos. The Kappa statistic was used to determine the reliability between the two raters. Inter-rater reliability was excellent (Kappa = 0.96, $P < .000$, $N = 150$).

We ran a generalised linear mixed model (GLMM) with delay condition as fixed and subject as random factor to determine whether the delay condition significantly affected the number of correct choices. Furthermore, we conducted one-sample t tests to determine in

which of the six delay conditions the subjects performed above chance levels (more than 50% correct choices) and whether their retest performance was still in the range of the criterion to which they had been trained in the pretest phase.

Ethical Note

This and the following experiment were conducted under guidelines established by the National Veterinary Office of Switzerland and licensed by the Veterinary Office of the Canton of Zurich (licence number 183/13).

3.3.3 Results

In the pretest phase, the marmosets reached criterion (≥ 80% correct choices within a single session) within 1 - 11 sessions of 10 trials each (*mean* = 2.93, *SD* = 2.55, t_{14} = 3.70, *P* = .002). In the test phase, the marmosets chose the correct container across delay conditions in 59% (*SD* = 8%) of all trials and thus significantly above chance (t_{14} = 4.04, *P* = .001). The GLMM with delay condition as fixed factor and subject as random factor showed that the duration of the delay had a significant effect on the subjects' performance ($F_{5, 63.77}$ = 3.31, *P* = .010). We had also predicted that the subjects' performance in the test phase of experiment 1 would decline with increasing length of the time delay, consistent with the Ebbinghaus effect. However, after an initial decline in the number of correct choices that was in line with this prediction, the subjects showed improved performance in the longest two delay conditions (Figure 3.1). A one-sample t-test demonstrated that the marmosets performed significantly above chance after delays of 5 s (*mean* = 66%, *SD* = 12%, t_{14} = 5.12, *P* < .001), 10 s (*mean* = 59%, *SD* = 15%, t_{14} = 2.42, *P* = .030) and 25 s (*mean* = 68%, *SD* = 16%, t_{11} = 4.01, *P* = .002), but not after delays of 15 s (*mean* = 48%, *SD* = 17%, t_{13} = -0.34, *P* = .741), 20 s (*mean* = 46%, *SD* = 16%, t_{12} = -0.81, *P* = .432) and 30 s (*mean* = 58%, *SD* = 17%, t_{11} = 1.70, *P* = .117).

In the retest no-delay condition, the marmosets chose the correct container in 66% of all trials (*SD* = 16%), which is significantly above chance (t_{10} = 3.46, *P* = .006) and higher than in five of the six test conditions, but differs significantly from the initial 80% criterion in the pretest (*mean* = 83%, *SD* = 5%; t_{10} = 3.33, *P* = .008).

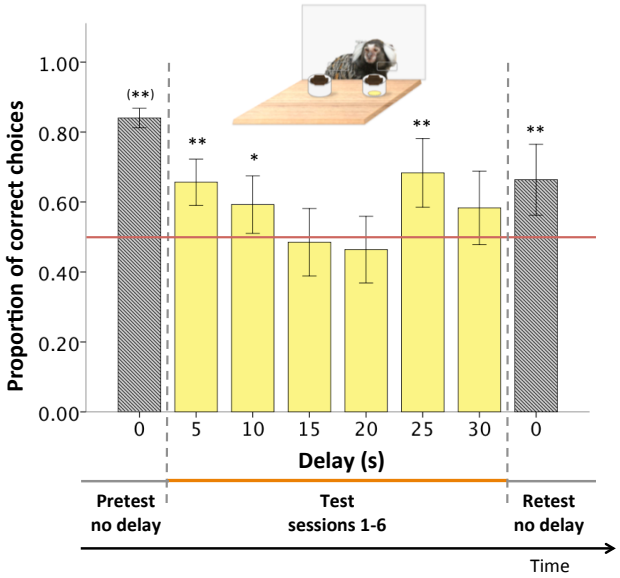


Figure 3.1 Performance in experiment 1. Subjects had to reach criterion (≥ 80% correct trials within a single session) in the pre-test phase (no delay) before entering the test phase (delays = 5 to 30 s) and were re-tested without a delay after completing the test phase. The red line indicates the chance level of 50%. Significance levels for above chance performance are indicated by $*P < .05$, $**P < .01$. Error bars: 95% confidence intervals. *Apparatus not drawn to scale.*

3.3.4 Discussion

We tested 15 common marmosets with a traditional memory test (Banerjee et al, 2009). In this memory test, the subjects had to remember, over several time delay intervals ranging from 5 to 30 s, in which of two locations the experimenter had hidden a reward. All subjects passed the pretest phase, in which no delay was imposed, and, as a group, the marmosets also passed the test phase, by overall performing above chance. In contrast to our predictions, however, the marmosets' performance in the memory test did not decline with increasing delay duration, and they showed low levels of correct performance. It is unlikely that the marmosets were unable to remember the reward's location since they performed well after relatively long delays of up to half a minute. Moreover, saddle-back tamarins, *Saguinus fuscicollis*, another callitrichid species and close phylogenetic relative, have been shown to remember the location of food items over much longer delay intervals of up to 24 h when tested in a naturalistic foraging task (Menzel, Juno and Garrod, 1985). An alternative explanation for the marmosets' unexpected performance in the hidden reward retrieval test (experiment 1) is that they may not have been sufficiently motivated to choose correctly, particularly after short delays, because of the low cost of a wrong choice. When choosing randomly between the two possible reward locations, they still had a 50% chance of receiving a reward in each trial, and it was only after longer delays between the experimenter's hiding

action and the subject's choice that the cost of a wrong choice increased owing to the longer waiting period.

We therefore designed a second experiment to explore whether the task format, i.e., the number of choice options, could explain the unexpected pattern of results in the traditional memory subtest. Based on the findings of Burkart & Heschl (2006) in a modified object choice task and our results from experiment 1, we developed a new memory test consisting of nine choice options. This reduced the probability of making a correct choice by chance from 50% in the two-choice memory test to 11% and thus made a subject's wrong choice more costly. We investigated whether this nine-choice format, which had been shown to increase the performance of marmosets in the above-mentioned social cognition task, would also be more suitable than the two-choice format in physical cognition tests. To do so, we compared the performance of a naïve marmoset group in the traditional and our new memory test. In addition, we also tested a group of squirrel monkeys with the same two task formats and directly compared the performance of the two species. This allowed us to evaluate whether task format effects are specific to common marmosets or also present in other nonhuman primates. We expected both species to perform better in the nine-choice memory test and to find an Ebbinghaus effect. Furthermore, we expected the squirrel monkeys to outperform the marmosets as they have in previous delayed response studies (Miles & Meyer, 1956; Miles, 1957b; Treichler, 1964; Tsujimoto & Savaguchi, 2002), owing to their larger absolute and relative brain size (in proportion to their small body size, squirrel monkeys have the largest brains of all primates; Rowe, 1996), which correlates with general performance in physical cognition tasks (Deaner, van Schaik, & Johnson, 2006; Reader, Hager, & Laland, 2011).

3.4 Experiment 2: Introducing a new memory test

3.4.1 Methods

Subjects

We tested 15 common marmosets with a traditional memory test (Banerjee et al., 2009). In this memory test, the subjects had to remember, over several time delay intervals ranging from 5 to 30 s, in which of two locations the experimenter had hidden a reward. All subjects passed the pretest phase, in which no delay was imposed, and, as a group, the

marmosets also passed the test phase, by overall performing above chance. In contrast to our predictions, however, the marmosets' performance in the memory test did not decline with increasing delay duration, and they showed low levels of correct performance. It is unlikely that the marmosets were unable to remember the reward's location since they performed well after relatively long delays of up to half a minute. Moreover, saddle-back tamarins, *Saguinus fuscicollis*, another callitrichid species and close phylogenetic relative, have been shown to remember the location of food items over much longer delay intervals of up to 24 h when tested in a naturalistic foraging task (Menzel, Juno, & Garrod, 1985). An alternative explanation for the marmosets' unexpected performance in the hidden reward retrieval test (experiment 1) is that they may not have been sufficiently motivated to choose correctly, particularly after short delays, because of the low cost of a wrong choice. When choosing randomly between the two possible reward locations, they still had a 50% chance of receiving a reward in each trial, and it was only after longer delays between the experimenter's hiding action and the subject's choice that the cost of a wrong choice increased owing to the longer waiting period.

We therefore designed a second experiment to explore whether the task format, i.e., the number of choice options, could explain the unexpected pattern of results in the traditional memory subtest. Based on the findings of Burkart & Heschl (2006) in a modified object choice task and our results from experiment 1, we developed a new memory test consisting of nine choice options. This reduced the probability of making a correct choice by chance from 50% in the two-choice memory test to 11% and thus made a subject's wrong choice more costly. We investigated whether this nine-choice format, which had been shown to increase the performance of marmosets in the above-mentioned social cognition task, would also be more suitable than the two-choice format in physical cognition tests. To do so, we compared the performance of a naïve marmoset group in the traditional and our new memory test. In addition, we also tested a group of squirrel monkeys with the same two task formats and directly compared the performance of the two species. This allowed us to evaluate whether task format effects are specific to common marmosets or also present in other nonhuman primates. We expected both species to perform better in the nine-choice memory test. Furthermore, we expected the squirrel monkeys to outperform the marmosets as they have in previous delayed response studies (Miles & Meyer, 1956; Miles, 1957b; Treichler, 1964; Tsujimoto & Savaguchi, 2002), owing to their larger absolute and relative brain size (in proportion to their small body size, squirrel monkeys have the largest brains of all primates;

Rowe, 1996), which correlates with general performance in physical cognition tasks (Deaner, van Schaik, & Johnson, 2006; Reader, Hager, & Laland, 2011).

Set-up

All subjects of both species were tested individually in a separate test compartment of a larger test enclosure. The measurements of the marmosets' test compartment closely resembled those in experiment 1, whereas the squirrel monkeys' test compartment measured 110 x 98 cm and 77 cm high. We again used a test apparatus that could be slid forwards and backwards on a test table (Figure 2). The apparatuses for the marmosets (M) and the squirrel monkeys (S) were identical and differed only in measurements that were adjusted to the marmosets' smaller body size. They consisted of a wooden frame (M: 40 cm x 37.5 cm; S: 80 cm x 75 cm) containing three wooden platforms (vertical distance between platforms: M: 12.5 cm; S: 35 cm) that was mounted on a wooden sliding board (M: 45 cm x 30 cm; S: 95 cm x 50 cm). Empty cylindrical black plastic cups (diameter: 3.1 cm; height: M: 1.1 cm; S: 2.3 cm) with lids were used to hide the reward. For the nine-choice test, three cups were placed equidistant (M: 14 cm; S: 29 cm) between each outer and the middle cup) on each platform (outer cups at M: 4.5 cm; S: 11 cm from the lateral frame). For the two-choice test, two cups were placed on the middle platform (in-between distance: M: 11 cm; S: 25 cm; distance from the lateral frames: M: 10 cm; S: 25 cm). In both tests, the cups were held in place by Velcro tape strips. The front of the test enclosure consisted of a lattice that allowed the subjects to reach out and choose one of the cups.

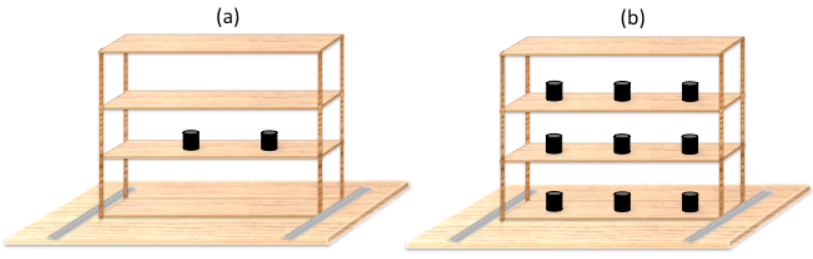


Figure 3.2 Test apparatus in experiment 2. Shown are both tests: (A) two-choice, and (B) nine-choice task format. *Not drawn to scale, the lateral parts of the frame were solid*

Procedure

The experimenter's procedure in the pretest and test phase corresponded to that used in experiment 1 with the exception of two additions in the test phase: (1) the experimenter said 'come' while pushing the apparatus towards the subject once the delay had expired in order to encourage the subject to make its choice, and (2) the subject received one to three warm-up trials (no delay) prior to each test session, and the test session only started once it had chosen correctly in a warm-up trial. There were four increasing delay conditions ranging from 5 to 20 s and each test session consisted of 12 trials of one delay condition. When choosing correctly, the subjects received their favourite rewards, a mealworm or a piece of a cashew nut (squirrel monkeys) and a cricket or a piece of a cooked apple (marmosets). The same stop criteria as in experiment 1 were used to decide when to terminate a session and continue testing on the next day.

We used a within-subject design in which every subject of each species was tested with both task formats, the one with two choice options and the one with nine choice options, in counterbalanced order. This resulted in two groups within each species: one group first completed the two-choice format followed by the nine-choice format while the second group was tested in the opposite order. One male marmoset (Lexus) completed the whole two-choice memory test but only the 5 s delay condition in the nine-choice memory test. The final sample size therefore consisted of eight marmosets (four females and four males) in the two-choice and seven marmosets (four females and three males) who completed all conditions in the nine-choice test, as well as seven male squirrel monkeys, who completed both tests.

3.4.2 Data scoring and analysis

All trials were video-recorded and the experimenter coded the subjects' choices live using check sheets. A second rater coded 21% of the trials from videos. The Kappa statistic was used to determine the reliability between the two raters. For the marmosets one trial had to be excluded owing to experimenter error (no delay imposed). Inter-rater reliability was excellent (100%) for both squirrel monkeys (Kappa = 1.00, $P < .001$, $N = 144$) and marmosets (Kappa = 1.00, $P < .001$, $N = 156$).

To test which factors best explained the subjects' performance, we ran GLMMs using the restricted maximum likelihood method, with the fixed factors task format, species, delay

condition, test order and interactions (species*task format, species*delay, species*order, task format*delay, task format*order, delay*order), and included subject as random factor. The best model was determined using the AICc, the Akaike information criterion corrected for small sample sizes (Hurvich & Tsai, 1989).

Since the probability to be successful by chance differed between the two-choice and nine-choice format, we could not use the subjects' raw scores to directly compare their performance in the two conditions in the same statistical model but first had to compute a performance measure that was independent of the task format. For this purpose, we computed a performance score for each individual and condition according to the equation below, which corresponds to the square root of the chi-square value and in which a higher value represents better performance. Observed values correspond to the individual number of correct choices per delay (raw scores of 1 - 12) and expected values were calculated as the number of correct choices expected by chance (six out of 12 in the two-choice and 1.33 out of 12 in the nine-choice memory test):

$$\text{Performance score} = \frac{(\text{observed} - \text{expected})}{\sqrt{\text{expected}}}$$

Finally, we conducted one-sample t tests for each test format to determine in which conditions the subjects performed above chance.

3.4.3 Results

In the pretest phase, the subjects reached criterion (≥ 80% correct within a single session) after one to two sessions (*mean* = 1.07, *SD* = 0.26) in the two-choice memory test and after one to seven sessions (*mean* = 2.27, *SD* = 1.71) in the nine-choice task. The subjects took significantly longer to reach criterion in the nine-choice than the two-choice task (t_{14} = -2.61, *P* = .021). There was no species difference in the number of pretest sessions in the two-choice (squirrel monkeys: *mean* = 1.00, *SD* = 0.00; marmosets: *mean* = 1.13, *SD* = 0.35; t_{13} = – 0.93, *P* = .369) or in the nine-choice memory test (squirrel monkeys: *mean* = 1.57, *SD* = 1.57; marmosets: *mean* = 2.88, *SD* = 2.10; t_{13} = -1.54, *P* = .015).

The best model included only the fixed-effects test format and delay condition and no interactions. Task format had a highly significant effect on the subjects' performance ($F_{1,\ 98.98}$ = 18.13, *P* < .001) and so did delay condition ($F_{1,\ 98.29}$ = 5.65, *P* = .001). There was no

significant effect of species in any of the models. Two separate GLMMs based on raw scores of performance, one for each task format, with species, delay and order as fixed factors and subject as random factor, demonstrated that delay condition had a significant effect on the subjects' percentage of correct choices for the nine-choice format ($F_{1, 39.88}$ = 5.46, P = .003) while there was only a trend for the two-choice format ($F_{1, 42}$ = 2.49, P = .073; see also Figure 3.3).

In the two-choice memory test, mean performance across all test sessions was significantly above chance, i.e., > 50% correct choices (*mean* = 59%, *SD* = 18%, t_{59} = 4.02, P < .001). Split per delay condition, the subjects as a group performed significantly above chance in the 5 s (*mean* = 61%, *SD* = 15%, t_{14} = 2.87, P = .015) and 10 s (*mean* = 65%, *SD* = 19%, t_{14} = 2.97, P = .010) delay conditions, but not in the 15 s (*mean* = 60%, *SD* = 19%, t_{14} = 1.92, P = .076) and 20 s conditions (*mean* = 52%, *SD* = 16%, t_{14} = 0.40, P = .695). In the nine-choice memory test, they also performed significantly above chance, i.e., > 11% correct choices, across all test sessions (*mean* = 25%, *SD* = 19%, t_{56} = 5.43, P < .001). Moreover, they performed well above chance after delays of 5 s (*mean* = 34%, *SD* = 21%, t_{14} = 4.13, P = .001), 10 s (*mean* = 26%, *SD* = 19%, t_{13} = 3.00, P = .010), 15 s (*mean* = 26%, *SD* = 20%, t_{13} = 2.79, P = .015), but not 20 s (*mean* = 14%, *SD* = 13%, t_{13} = 0.786, P = .446).

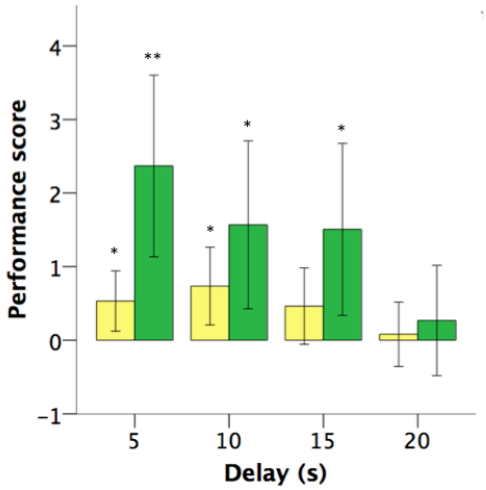


Figure 3.3 The effect of task format on performance in experiment 2. The subjects' mean performance (χ-transformed test scores to account for the two different chance levels) in the two-choice (yellow/light bars) and the nine-choice memory test (green/dark bars). Error bars: 95% confidence interval. Asterisks * and ** indicate performance significantly above chance (P < .05 and P < .01, respectively) in one-sample-t-tests on the raw values (per cent correct choices).

3.4.4 Discussion

In experiment 2, we tested common marmosets and common squirrel monkeys, two evolutionarily closely related species, with both the two-choice and nine-choice task format of a memory test. As predicted, we found that task format affected the performance of both species. When the subjects were allowed to choose between nine rather than only two options, they performed better, and, in line with our prediction, their performance decreased with longer delays, consistent with the Ebbinghaus effect. However, the larger-brained squirrel monkeys did not outperform the marmosets, although the small sample size makes it difficult to identify whether this finding is a true absence of a species difference or instead reflects a lack of statistical power. In contrast to the present findings, squirrel monkeys had outcompeted marmosets in delayed response studies. However, some of these studies did not contain a true memory component (Miles & Meyer, 1956; Miles, 1957b) or they contained a stronger working memory component (Tsujimoto & Savaguchi, 2002). In the latter study, the reward's location was not randomised and subjects had to keep in mind their previous choices and base their next choices strategically on these.

Both New World monkey species in our study remembered the reward's location for longer time periods in the nine-choice format than in the two-choice format. However, their performance was still moderate, for instance in relation to closely related saddle-back tamarins (Menzel et al., 1985) who remembered as many as 30 different locations over much longer delays. Apart from species differences, a likely explanation is that the tamarin study was more naturalistic than ours in that the subjects remained in the group setting during experiments, and the locations to choose from were distributed over a much larger area. In fact, Fischer and Kitchener (1965) had argued that delayed-response tasks with a strong spatial component are easier to solve for nonhuman primates than nonspatial ones. The more pronounced spatial component may thus have tapped into the tamarins' adaptive behaviour as extractive foragers (Peres, 1992) and thereby made the tasks easier to solve. However, whenever the aim is to assess an individual's cognitive ability per se, rather than a specific adaptation to a narrow cognitive problem, it is preferable to present a task in an abstract rather than a naturalistic manner. This is perhaps most evident for general intelligence test batteries that consist of a number of subtests assessing a wide range of abilities from various cognitive domains to identify whether they are all based on a single domain-general cognitive ability (Burkart, Schubiger & van Schaik, 2016). The traditional two-choice memory test in experiment 1 is a

subtest of one such test battery, and we developed our nine-choice memory test as a possible alternative.

3.5 General discussion

We conducted two experiments in order to explore whether the task format affects cognitive performance of nonhuman primates in physical cognition tests as has been reported for a widely used social cognition test (Burkart & Heschl, 2006). When testing marmosets with a traditional two-choice memory test (experiment 1), we found that, in contrast to the Ebbinghaus effect, their performance did not continuously decline with increasing delay duration. To address the possibility that our results reflected a lack of motivation to memorise the location of the food rather than the marmosets' ability to do so, we designed a new version of the memory test (experiment 2) with nine choice options instead of two, which lowered the probability of making a correct choice by chance from 50% to 11%. Both marmosets and squirrel monkeys performed better in the nine-choice memory test, and their performance now continuously decreased with increasing delay duration, consistent with the Ebbinghaus effect we had predicted. Our results suggest that the nine-choice format is more accurate in assessing memory performance in the two New World monkey species, and that the two-choice format negatively affects performance not only in a social cognition task, but also in a physical one.

Our findings have important implications for studies that assess cognitive performance in nonhuman primates and other animals for comparative purposes. Examples of such comparisons include the assessment of differences in cognitive performance across different tasks between individuals of one species (e.g. to investigate general intelligence; Banerjee et al., 2009; Herrmann, Hernàndez-Lloreda, Call, Hare & Tomasello, 2010), between conspecifics differing in certain traits (e.g., to investigate sex differences; Schubiger et al., 2015) or environmental/ontogenetic conditions (e.g., to investigate rearing differences; Damerius & Forss et al., 2016; Herrmann & Call, 2012), and differences in cognitive performance between species (i.e., to investigate evolutionary trajectories; Amici, Aureli, & Call, 2008, 2010). For all these comparative purposes it is crucial that differences in measured performance reflect true differences in the subjects' cognitive abilities and cannot be attributed to differences in their motivation to engage with a specific task.

Decreasing the chance level probability of success, as we have done in the present study, is one way of promoting the subjects' motivation. But although using more than two choice options is advantageous in some cognitive tests with animals, it is probably not feasible in others. Examples of physical cognition tests that require the two-choice format are those in which the subject has to base its choice on more or less apparent differences in the amount (e.g., numerical discrimination tests, Agrillo, 2014) or external features (e.g., tool functionality, Mulcahy & Schubiger, 2014) of the test stimuli. In such tests, additional options could either lead to ambiguous choices or be too demanding for a subject's working memory. However, the costs of a wrong choice can also be increased in two-choice tests, e.g., by requiring subjects to choose by performing an effortful behavioural response such as unscrewing a lid, pulling in the chosen item, or a similarly effortful behaviour.

In sum, we found that nonhuman primates may not be sufficiently motivated to fully engage in a cognitive task when presented in a two-choice format but that some methodological modifications can restore their motivation.

If future studies show that our findings generalize to other species beyond marmosets and squirrel monkeys, and to cognitive domains other than memory, it may be preferable to replace the two-choice format with alternative task formats. Otherwise, cognitive performance may be biased in comparisons both within and across species, for instance towards more food-motivated individuals or species.

3.6 Acknowledgements

We thank Carel van Schaik and Nick Mulcahy for discussion, Erik Willems for statistical advice, head animal keeper, Heinz Galli, for help with building the experimental apparatuses, and animal keepers, Thomas Bischof and Patricia Rivera, for taking good care of the monkeys. We also thank two anonymous referees for their helpful comments. This study was financially supported by the Swiss National Science Foundation (project number 310030-130383) and two personal grants to Michèle Schubiger by the A. H. Schultz-Foundation.

Chapter 4

Does opportunistic testing bias cognitive performance in primates? Learning from drop-outs

Michèle N. Schubiger[1,2*], Alexandra Kissling[1], & Judith M. Burkart[1]

[1]Department of Anthropology, Evolutionary Cognition Group, University of Zurich, Zurich, Switzerland

[2]School of Social and Health Sciences, Division of Psychology, Abertay University, Dundee, Angus, United Kingdom

Under review:

Submitted to *PLoS One*: 17th March 2018; resubmitted: 31st July 2018.

4.1 Abstract

Dropouts are a common issue in cognitive tests with non-human primates. One main reason for dropouts is that researchers often face a trade-off between obtaining a sufficiently large sample size and logistic restrictions, such as limited access to testing facilities. The commonly used opportunistic testing approach deals with this trade-off by only testing those individuals who readily participate and complete the cognitive tasks within a given time frame. All other individuals are excluded from further testing and data analysis. However, it is unknown if this approach merely excludes subjects who are not consistently motivated to participate, or if these dropouts systematically differ in cognitive ability. If the latter holds, the selection bias resulting from opportunistic testing would systematically affect performance scores and thus comparisons between individuals and species. We assessed the potential effects of opportunistic testing on cognitive performance in common marmosets (*Callithrix jacchus)* and squirrel monkeys *(Saimiri sciureus)* with a test battery consisting of six cognitive tests: two inhibition tasks (Detour Reaching and A-not-B), one cognitive flexibility task (Reversal Learning), one quantity discrimination task, and two memory tasks. Importantly, we used a full testing approach in which subjects were given as much time as they required to complete each task. For each task, we then compared the performance of subjects who completed the task within the expected number of testing days with those subjects who needed more testing time. We found that the two groups did not differ in task performance, and therefore opportunistic testing would have been justified without risking biased results. If our findings generalise to other species, maximising sample sizes by only testing consistently motivated subjects will be a valid alternative whenever full testing is not feasible.

4.2 Introduction

Cognitive performance in animals can be assessed as pure proof of principle, i.e. to test whether a specific ability can be found in a given species. However, cognitive performance is increasingly assessed also for comparative purposes, both between and within species. At the inter-specific level, two or more species are tested with the same cognitive task(s) in order to explore species differences and similarities in cognitive abilities. These differences are usually hypothesised to emerge because the species have faced different selection pressures in their evolutionary past, such as challenges posed by their social or ecological environment. Examples of such studies suggest that enhanced inhibitory control is found in species with a fission-fusion social group structure (Amici, Aureli, & Call, 2008) or a feeding ecology that requires more patience (Stevens, Hallinan, & Hauser, 2005), that more tolerant species show increased socio-cognitive performance (Joly et al., 2017), and that species who rely on a diet rich in fruit (Rosati, Rodrigues, & Hare, 2014) or food caching species (Prasvosudov & Roth, 2013) show enhanced spatial memory. Encountering specific social and ecological challenges can thus, over evolutionary time, lead to domain-specific cognitive adaptations.

Cognitive comparisons can also focus on the individual level, allowing researchers to investigate differences and similarities between individuals of the same species. Examples include studies exploring the effects of aging (Lacreuse et al., 2014; Gray et al., 2017; Mattison & Vaughan, 2017), sex differences (e.g., Wobber et al., 2014; Schubiger et al., 2015), or links between cognitive performance and personality (e.g., Carere & Locurto, 2011; Griffin, Guiliette, & Healy, 2015). Finally, individual differences can be compared across a variety of cognitive tasks. If cognitive abilities co-vary across individuals of the same species, this is consistent with the notion of domain-general cognition rather than domain-specific cognitive adaptations, and thus general intelligence which facilitates solving a wide range of problems, particularly novel ones (see Burkart, Schubiger, & van Schaik, 2017 for a review).

Whenever the goal is to compare cognitive performance across species and individuals, it is crucial to take into account that an individual's performance in a test is also affected by factors that are not primarily cognitive in nature, such as emotion, motivation and health. We may thus risk measuring individual differences in such non-cognitive factors rather than cognitive ability per se and therefore the performance scores would be biased or even

meaningless. The same is true if individuals are selected for inclusion in a cognitive study based on such non-cognitive factors.

In the last decade, researchers have started to systematically re-address the risk that cognitive tests may inadvertently measure individual (and species) differences in non-cognitive factors rather than true differences in the cognitive abilities. A number of experimental studies in non-human primates has systematically assessed both external (testing-related) and internal (subject-related) factors to determine how they affect the subjects' cognitive performance (Table 1). Examples for internal factors that affect cognitive performance in physical cognition tasks include individual differences in the subjects' psychological predispositions (e.g., temperament and personality structure (Toxopeus et al., 2005; Herrmann et al., 2007; Morton, Lee, & Buchanan-Smith, 2013) , and emotional reactivity (Schubiger et al., 2015) and individual differences in the subjects' ontogenetic experience with the social environment (e.g., rearing conditions (van Schaik & Burkart, 2011), level of contact with humans (Damerius et al., 2017a) and conspecifics (Damerius et al., 2017b).

Examples of external non-cognitive factors that have been addressed so far include aspects of the test design, setup and procedure such as how the test apparatus and test stimuli are presented to the subject (e.g., task format (Schmitt & Fischer, 2011; Gazes, Billas, & Schmidt, 2018; Schubiger, Kissling, & Burkart, 2016; Carducci et al., 2018; also see Burkart & Heschl, 2006 for a similar effect in a social cognition task), and how subjects are rewarded when they pass a test trial (e.g., reward type (Gazes et al., 2018)).

Table 4.1 Non-cognitive internal (subject-related) and external (testing-related) factors potentially affecting cognitive performance (in physical cognition tasks). Increase (↑), decrease (↓) or no effect (=) on performance.

Factor	Cognitive task(s)/skills	Effect on performance?	Species	Reference
Trait anxiety	Reversal learning	Performance ↓ in subjects with trait anxiety	Long-tailed macaques (*Macaca fascicularis*)	Toxopeius et al., 2005
Temperament	Physical cognition	Performance ↑ in bolder subjects	Chimpanzees (*Pan troglodytes*) & orangutans (*Pongo pygmaeus*)	Herrmann et al., 2007
Personality	Training	Performance ↑ in subjects with high openness & low assertiveness scores	Brown capuchins (*Sapajus apella*)	Morton et al., 2013
Emotional reactivity	Object permanence	Participation ↓ but performance = in subjects with elevated arousal levels during testing	Common marmosets (*Callithrix jacchus*)	Schubiger et al., 2015

Factor	Cognitive task(s)/skills	Effect on performance?	Species	Reference
Rearing conditions	Repertoire of learned cognitive skills	(1) Skill repertoire ↑ in mother-reared individuals (2) Skill repertoire ↓ in orphaned Individuals	Various primate species	Reviewed in: van Schaik & Burkart, 2011
		Set of skills & learning speed ↑ in enculturated individuals; even beyond a species' natural repertoire	Great apes	
Degree of orientation towards humans	Problem-solving	Performance ↑ in subjects with high HOI (Human-Orientation Index)	Sumatran (*Pongo abelii*) & Bornean orangutans (*Pongo pygmaeus*)	Damerius, Forss et al., 2017a
Human care & social housing with conspecifics	Inhibitory control, reversal learning, problem solving, causal reasoning	Performance ↑ with curiosity & exploration ("curious response-and-exploration style")		Damerius et al., 2017b
Task format	Quantity discrimination	Performance ↑ (1) with inedible test stimuli (2) with edible test stimuli if reward items differ in food type	Olive baboons (*Papio anubis*) & long-tailed macaques (*Macaca fascicularis*)	Schmitt et al., 2011
		Performance ↑ with edible test stimuli	Brown capuchins (*Cebus sapajus apella*)	Gazes et al., 2018
		Performance ↑ with rewards of higher value	Brown capuchins (*Cebus sapajus apella*) & common squirrel monkeys (*Saimiri sciureus*)	
	Memory	(1) Performance ↑ with more choice options (9 instead of 2) (2) Performance ↓ with delay duration (in line with forgetting curve)	Common squirrel monkeys (*Saimiri sciureus*) & common marmosets (*Callithrix jacchus*)	Schubiger et al., 2016[1]
	Visual object discrimination	Performance ↑ when tactile exploration of the objects is possible	Capuchin monkeys (*Sapajus spp.*)	Carducci et al., 2018
Opportunistic testing	Inhibitory control & memory	Performance = when excluding subjects who take longer to complete all test trials	Common marmosets (*Callithrix jacchus*) & common squirrel monkeys (*Saimiri sciureus*)	This study
	Quantity discrimination & reversal learning		Common marmosets (*Callithrix jacchus*)	

[1]see Burkart & Heschel, 2006 for a similar positive effect of more choice options in an object choice task

Despite these recent efforts to assess potential biases on cognitive performance, one important aspect of comparative testing has largely been neglected. Comparative testing requires a sufficiently large sample size to make valid inferences. Researchers therefore should test a sufficient number of individuals to reach adequate statistical power, often under logistic restrictions such as limited access to testing facilities. One way to deal with these constraints

is opportunistic testing, i.e. only testing those individuals who readily participate and respond to the experimental tasks and drop those who respond too slowly or to erratically. By only including the ready participants, researchers can maximise sample size if enough subjects but only limited testing time is available. Some subjects may be more reluctant to participate because of non-cognitive attributes such as a tendency to react to the test situation with elevated emotional arousal. Those subjects may no longer be motivated to participate in a cognitive test after a relatively low number of completed trials or get easily distracted during testing (as found in marmosets, particularly males (Schubiger et al., 2015). Other subjects, however, may be reluctant to participate because their cognitive skills do not allow them to pass the task and earn rewards. If so, their exclusion would be detrimental to any meaningful comparison of cognitive ability.

An open question is therefore whether opportunistic testing, here defined as using strict stop criteria to restrict total testing time, results in a selection bias because the subjects included in the study differ in cognitive ability from a random sample of the population.

The aim of the present study was to address this question in two New World primate species, common marmosets (*Callithrix jacchus*) and squirrel monkeys (*Saimiri sciureus*). The subjects were tested with a cognitive test battery consisting of five commonly used paradigms: Detour Reaching, A-not-B, Quantity Discrimination, Reversal Learning, and a two-choice Memory task (largely adapted after an existing test battery for New World monkeys (Banerjee et al., 2009), as well as a nine-choice Memory task (Schubiger, Kissling, & Burkart, 2016). Importantly, in contrast to opportunistic testing, we gave each subject as much time as it needed to complete each task at its own pace. We then compared, for each cognitive task, the performance of subjects who had completed the task within the expected time frame (i.e., a predefined number of days or trials to complete the task) to the performance of those subjects who needed longer. If cognitive ability had an effect on how long an individual took to complete a cognitive test, then subjects who needed longer to complete a test should either perform worse or better than those subjects who completed the test in the expected time frame. In this case, opportunistic testing would indeed bias the results and would need to be abandoned in future studies. If there was no difference between the two groups, however, the commonly used opportunistic testing approach that excludes some subjects would be equally justified without risking a selection bias in the results.

4.3 Materials and methods

4.3.1 Study sample

As summarized in Table 4.2, 27 adult common marmosets (*Callithrix jacchus*) from six family groups and 8 adult common squirrel monkeys (*Saimiri sciureus*) from two bachelor groups participated in this study. The initial study sample consisted of the 15 marmosets (7 females and 8 males) in the upper section of the table who were tested with the whole test battery of subtests 1-5 (hereafter tasks). Kitty (female) who became Kantor's partner later in the study (after he had to be separated from his brothers Kaliper and Kapi) was tested with task 5. The remaining 11 marmosets (5 females and 6 males) were only tested with both memory tasks (5 and 6) when we assessed in a parallel study whether memory task 5 truly measured the ability to remember the location of a reward (Schubiger et al., 2016). The squirrel monkeys (8 males) were only tested with tasks 1, 2, 5 and 6 because their temporal availability for testing was restricted. The study subjects had previously participated in some cognitive experiments, but we randomly chose those subjects to include in our study who were available because they were not participating in other studies of cognitive or behavioural nature at the time. We tested all individuals of both squirrel monkey groups (the only two groups of this species that were housed at the Primate Station) and all adult individuals of the available marmoset family groups.

All monkeys were captive-born and mother-reared and housed in family (marmosets) and bachelor groups (squirrel monkeys) consisting of two to eight individuals in indoor-and outdoor enclosures. The marmosets' indoor enclosures had both daylight and artificial light and were composed of one to three components (depending on group size) measuring at least 4 m^3, each of which was equipped with several climbing structures such as natural branches, a sleeping box, an infrared lamp and a bark mulch floor. Whenever the weather conditions allowed it, each group had free access to an outdoor enclosure. The squirrel monkeys' indoor enclosures measured 16.55 m^3 (smaller group) and 24.77 m^3 (larger group) and were equipped with climbing structures, an infrared lamp and a bark mulch floor. Since the squirrel monkeys' indoor enclosures mainly had artificial UV-light, each group had constant free access to a fully roofed outdoor enclosure, and in addition, the two groups took turns in accessing a larger outdoor area of 86.4 m^3. The marmosets were fed a vitamin and calcium-enriched porridge in

the morning, fresh fruit and vegetables at lunchtime, and gum and mealworms in the late afternoon. In addition, they received a daily protein snack in the afternoon such as pieces of cooked egg. The squirrel monkeys were fed a mixture of pellets and cottage cheese in the morning, a variety of vegetables and a small amount of fruit at lunchtime, and a protein snack such as cockroaches in the late afternoon. Water was available ad libitum from water dispensers (marmosets) or water bowls (squirrel monkeys). All subjects were tested between their regular feeds and never food deprived during the study. Testing time was restricted to 8.30-12.00 hours in the morning and 13.00-16.30 in the afternoon. Each subject was only tested once per day and never during the monkeys' general resting time from 12.00-13.00 hours. The monkeys could freely enter and leave the test enclosure and were not handled by the experimenter at any time before, during or after the cognitive test sessions. After the completion of this study, the monkeys continued living at the Primate Station, eventually participating in other non-invasive studies.

Table 4.2 Individual subjects by species and family group, their sex (female/male), mean age (in years) during testing, time needed to complete and performance scores in each task. A performance score value in regular font indicates that the subject completed the cognitive task and its performance could be fully analysed whereas a performance score value in italics indicates that the subject did not complete the full task but its performance could be analysed for parts of the tasks. A dash (-) indicates that the subject participated in such a small number of trials that its performance could not be analysed and an empty cell indicates that a subject refused to participate at all. Grey cells indicate that a subject was not tested with a given task. (*See next page*).

ID	Individual	Species	Sex	Age	1. Detour Reaching			2. A-not-B		3. Quantity Discrimination		4. Reversal Learning		5. Memory 1				6. Memory 2	
					Days	Correct inhibition trials at 1st attempt	Total correct inhibition trials (incl. after initial failure)	B-trials to criterion	A-trial correct 1 = yes 0 = no	Days	Total correct trials	Sessions per day in pre-reversal	TI	Days (2nd value = 4 delay condition)	Total correct trials			Days	Total correct trials
															a	b	a & b combined		
M1	Juri	*C. j.*	m	11.2	6	0.10	1.00	5	0	-	-	0.5	0.22	21/13	0.60		0.58		
M2	Venezia	*C. j.*	f	6.6	5	0.50	0.90	12	0	3	0.67	0.9	0.00	12/7	0.61		0.62		
M3	Venus	*C. j.*	f	8.0	5	0.45	0.95	5	1	3	0.70	1	0.77	21/13	0.53		0.44		
M4	Verona	*C. j.*	f	7.1	5	0.25	0.90	11	1	3	0.67	1	0.44	10/6	0.57		0.53		
M5	Vesta	*C. j.*	f	7.9	5	0.25	1.00	7	1	3	0.80	1	0.22	8/7	0.57		0.53		
M6	Vito	*C. j.*	m	6.8	6	0.05	0.05	5	0	6	0.73	0.6	0.33	25/18	0.66		0.56		
M7	Vreni	*C. j.*	f	10.4	5	0.35	0.70	5	1	3	0.77	1	0.00						
M8	Jugo	*C. j.*	m	5.9	7	0.05	0.70	10	0	3	0.77	0.4	0.10	16/8	0.42		0.43		
M9	Tabor	*C. j.*	m	4.7	6	0.55	0.95	5	1	2	0.70	1	0.00	14/10	0.53		0.55		
M10	Tale	*C. j.*	f	3.8	10	0.30	0.80	5	1	3	0.83	1	0.20	8/6	0.60		0.55		
M11	Tessy	*C. j.*	f	11.6	5	0.00	0.00	7	1	3	0.50	0.9	0.49	12/6	0.50		0.45		
M12	Thilo	*C. j.*	m	5.0	11	0.50	0.85	5	1	4	0.67	1	0.22	12/6	0.68		0.70		
M13	Kaliper	*C. j.*	m	11.4	6	0.25	0.95	47	0	-	-	-	-	16/14	0.58		0.56		
M14	Kapi	*C. j.*	m	9.0	6	0.45	1.00	15	0	4	0.70	-	-	-/10	-		0.51		
M15	Kantor	*C. j.*	m	10.7	6	0.20	0.50	7	1	-	-	-	-	-/-	-		-		
M16	Kitty	*C. j.*	f	4.6										-/-	-		-		
M17	Lex	*C. j.*	m	8.8										4		0.69	0.69	8	0.13
M18	Nando	*C. j.*	m	2.0										11		0.60	0.60	5	0.04
M19	Nautilus	*C. j.*	m	2.5										9		0.48	0.48	12	0.10
M20	Nebula	*C. j.*	F	2.5										12		0.54	0.54	9	0.21
M21	Nina	*C. j.*	F	8.8										4		0.53	0.53	4	0.46
M22	Nuno	*C. j.*	m	2.0										-		-	-	-	-
M23	Lancia	*C. j.*	f	12.8										4		0.69	0.69	4	0.21
M24	Lexus	*C. j.*	m	12.8										8		0.56	0.56	-	-
M25	Lili	*C. j.*	f	2.0										4		0.58	0.58	4	0.21
M26	Lola	*C. j.*	f	2.3										-		-	-	-	-
M27	Lotus	*C. j.*	m	3.0										-		-	-	-	-
S1	Chipo	*S. s.*	m	9.6	5	0.00	0.00	5	1					4		0.90	0.90	4	0.58
S2	Chris	*S. s.*	m	7.6	5	0.00	0.05	5	1					4		0.60	0.60	4	0.10
S3	Darwin	*S. s.*	m	7.4	6	0.00	0.00	5	1					5		0.56	0.56	6	0.17
S4	Dave	*S. s.*	m	8.4	6	0.00	0.00	5	1					4		0.56	0.56	4	0.08
S5	George	*S. s.*	m	6.3	5	0.00	0.00	5	1					4		0.50	0.50	4	0.17
S6	Helio	*S. s.*	m	5.3	5	0.00	0.00	-	-					-		-	-	-	-
S7	Hugo	*S. s.*	m	2.1	6	0.65	1.00	5	1					4		0.79	0.79	4	0.40
S8	Iramo	*S. s.*	m	4.2	5	0.60	1.00	10	1					4		0.42	0.42	4	0.23

4.3.2 General set-up

All subjects were tested individually in a dedicated test enclosure that was connected to their home enclosures through a semi-transparent plastic tube. Once the doors on each end of the tube were opened, the subject could voluntary and freely enter the tube and walk to the test enclosure without being handled by humans at any time. The actual test compartment for the marmosets consisted of a white rear wall, a white floor, two lateral grid walls and a clear Perspex window front. Depending on the cognitive task and respective apparatus, the window front contained one or two rectangular openings through which the subject could reach with its arms to access the apparatus. Each test apparatus was placed on a wooden board (varying size and features) that was mounted on a height-adjustable test table and flush with the test compartment's window. The test compartment used for the squirrel monkeys also had a white plastic wall in the back and a white floor, and three grid walls. The measurements of the test apparatuses (Table S4.1) were adapted to the squirrel monkeys' larger body size (Rowe, 1996).

4.3.3 General procedure

The general testing procedure was identical for both species. Before and after each test trial, the experimenter positioned an occluder (a wooden or cardboard sheet) between the test compartment's window and the test apparatus. Its removal announced the next trial to the subject and the experimenter called the subject's name and said 'look' and made sure that the subject was attentive to the test apparatus and procedure. In the first task (Detour-Reaching), the test apparatus remained within the subject's reach and was accessible to the subject as soon as the experimenter had removed the occluder between the test compartment's front and the test apparatus. In all other tasks, the apparatus was mounted on a sliding platform and remained out of the subject's reach until the baiting process was completed (tasks 2-6) or the time delay had expired (tasks 5-6). The experimenter then pushed the test apparatus into the subject's reach so that the subject could make its choice. After a correct choice, the experimenter waited for the subject to consume the reward, retracted the test apparatus, set-up the occluder, and immediately continued with the next trial. After a wrong choice, however, the experimenter quickly retracted the test apparatus out of the subject's reach so that no second choice was possible. She then retrieved the reward, showed it to the subject and placed it in a holding area for re-use in the next test trial. In tasks 1 and 2, the next trial was started

immediately after the subject had completed a given test trial. In task 1, this ensured the monkeys' wellbeing by keeping the duration of each test session as short as possible because a single test trial took up to 2 minutes. In task 2, it was important to quickly continue with the next trial in order to get the subjects into the intended routine of repeatedly choosing the same cup. In tasks 3, 4, 5 and 6, the experimenter only continued with the next test trial once a 15-second (tasks 3-4) or 10-second (tasks 5-6) time interval had expired. This brief delay was intended to make a wrong choice more costly and encourage subjects to concentrate and attempt to choose correctly. Again, since test sessions in tasks 5 and 6 took inherently longer than in the other tasks because each test trial already involved a time delay of up to 30 seconds, the slightly shorter additional inter-trial interval after a wrong choice kept total testing time to a minimum. It is unlikely that this slight difference in inter-trial intervals was perceived by or affected the subjects' motivation differently because in each task a beep tone announced the end of the delay after a wrong choice. Since food preferences differed between the two species, marmosets received mealworms, crickets or locusts as rewards, whereas squirrel monkeys received mealworms or small pieces of cashew nut.

If a subject lost motivation during a test session and expressed this by refusing to make a choice in 3 consecutive test trials or showed behavioural signs of emotional arousal or avoidance (as determined by the stop criteria in Schubiger et al., 2015), the experimenter stopped testing and continued the test session on the following day. Importantly, we gave each subject as much time as it needed to complete each task during a total testing period of one to several months.

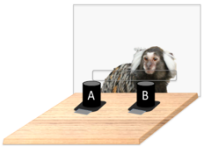
The criteria to determine the expected amount of testing time in which each task would have had to be completed under an opportunistic testing regime were adopted from a similar study (Banerjee et al., 2009) with the exception of tasks 4 and 5a for which these stop criteria would have been too strict. For tasks 4 and 5a, we defined expected testing time as the amount of time in which the majority of subjects had completed the task, because applying the above-mentioned stop criteria would have resulted in most or all subjects being excluded from further testing.

For those subjects who needed longer than expected to complete a task but were not excluded under our full testing regime, we determined the maximum total testing period for each task based on what appeared reasonable depending on the total number of test trials (one to five months). This extended time period allowed subjects ample opportunity to

complete each task at its own pace including breaks from testing as required (several days or weeks). We discontinued testing after this time period had expired because it was not justified to continue testing for a completely unlimited time period for animal welfare reasons.

In this section, we provide an overview of the 6 subtests of the test battery (see also Table 4.3). Further details regarding the test apparatuses are available in Table S4.1.

Table 4.3 Overview of the 6 tasks of the cognitive test battery.

Task	Measured cognitive ability	Procedure, Sessions, Trials	Performance measure(s)
1. Detour-Reaching 	**Inhibition of a pre-potent (motor) response** to directly reach for the reward in central trials	Reward in one of 3 positions behind Perspex panel; 5x12 = 60 trials: 20 central (fully behind panel) 20 left (half exposed) 20 right (half exposed)	**Successful detour-reaching**: % correct of 20 inhibition (central) trials at 1st attempt **Learning from error**: % correct of 20 inhibition trials after initial failure
2. A-not-B 	**Inhibition of a learned routine** to again choose cup B in the last trial in which the reward is placed under cup A instead	Reward is placed under one of two cups (A and B); (1) 5x (or more) under cup B until correct choice in five consecutive trials (2) 1x under cup A in last trial	**Correct choice** on last trial: cup A yes/no
3. Quantity Discrimination 	**Understanding of quantities** by choosing the larger of two amounts of food pellets	Simultaneous presentation of two amounts of food pellets; all possible pairs of amounts 1-5; 3x10 = 30 trials	**Choice of larger quantity**: % correct of all 30 trials
4. Reversal Learning 	**Associative learning & reversal** 1. Learning of a pattern-reward association 2. Reinforcement of the learned association 3. Reversal of learned the association	Simultaneous presentation of two patterns in 3 test phases: 1. Food visible; then covered by pattern 2. Food hidden; only patterns visible 3. Food hidden; but now under opposite pattern Number of 12-trial sessions to reach criterion in each phase	**Associative Learning**: Number of sessions to learn association in phases 1 & 2 **Complete Reversal**: Number of sessions to learn reversed association in phase 3 **Transfer Index**[1]: Reversal (% correct trials 2-12 in 1st session of phase 3)/pre-reversal (% correct trials in last session of phase 2)
5. Memory 1 	**Remembering the reward's location** after a time delay (seconds): (a) 5, 10, 15, 20, 25, 30 (b) 5, 10, 15, 20 (a & b) 5, 10, 15, 20	Reward is placed in one of two possible locations: (a) 6 conditions: 6x10 = 60 trials (b) 4 conditions: 4x12 = 48 trials (a & b) 4 conditions: 4x12 = 48 trials	**Remembering one of two locations**: (a) % correct choices across all 6 delays (sample 1) (b) % correct choices across all 4 delays (sample 2) (a & b) % correct choices across all 4 delays (sample 1 & 2)
6. Memory 2 	**Remembering the reward's location** after a time delay (seconds): 5, 10, 15, 20	Reward is placed in one of nine possible locations: 4x12 = 48 trials	**Remembering one of nine locations:** % correct choices across all 4 delays

[1](Rumbaugh, 1997)

4.3.4 The cognitive test battery

Task 1: Inhibition a – Detour Reaching

Measured cognitive ability. Inhibition of a predominant motor response (directly reaching for a food item that is placed behind a transparent barrier) in favour of applying an adequate and successful response (reaching around the barrier to be able to successfully access the food item).

Test apparatus. A static transparent barrier (a quadratic Perspex panel) was vertically attached to the top of a wooden board. The task required the subject to reach around the barrier in order to grasp a food item that was placed behind it.

Familiarisation phase. The subject was given the opportunity to explore the test apparatus in the test compartment without a food reward present for a duration of 10 minutes.

Test phase. In each trial, the experimenter placed the food item on the board and behind the Perspex barrier in one of 3 positions: left, central or right. In the two lateral conditions that simply served as distractors, only half of the reward was positioned behind the left or right-side edge of the Perspex barrier while its other half remained exposed. To successfully retrieve the reward, the subject could directly reach for it and grasp it. In the actual test condition (central trials), however, the reward was placed in the middle of the Perspex barrier so that it was fully occluded by it. To successfully retrieve the reward, the subject had to inhibit directly reaching for the reward as this would have resulted in its hand colliding with the Perspex barrier. It had to reach around the barrier instead so that it could grasp the reward without reaching into the barrier (successful Detour Reaching).

If the subject successfully reached around the barrier at its first attempt, the trial was noted as correct. If it failed to do so and reached into the barrier instead, the trial was noted as incorrect. Yet, contrarily to an earlier study (Banerjee et al., 2009), the apparatus was kept in place after such initial failure for a maximum of 2 minutes, which allowed the subject the opportunity to make further attempts to reach around the barrier. If a subject eventually succeeded within the 2-minute period, the trial was noted as correct after initial failure and the next trial was started. If the subject did not make any further detour- reaching attempts or still failed to successfully reach around the barrier, the next trial was started after the 2 minutes had expired.

The Detour-Reaching task entailed 5 sessions of 12 trials. In each test session, the reward appeared 4 times in each of the 3 positions (left, central, right) in a counterbalanced and pseudo-randomised order, with the rule that the reward never appeared in the same location on more than two consecutive trials. This generally resulted in the completion of the full 5 test sessions and 60 trials in 5 consecutive days. However, subjects who did not complete a whole session or refused to participate at all on a given day were tested the following day. This resulted in the test sessions being spread over more than the expected 5 days. We stopped a test session on a given day if a subject was not motivated to complete all trials. However, we allowed these subjects to eventually complete all 5 sessions of the task in as many testing days as it needed.

Outcome variables. (1) Successful detour-reaching: per cent correct trials of the 20 central trials at first attempt. (2) Learning from error: per cent correct trials at first attempt plus after initial failure (further detour-reaching attempts).

Stop criterion in opportunistic testing. In opportunistic testing, the five sessions of the Detour-Reaching task would have been supposed to be completed within five consecutive days (see Occluded Reach task, (Banerjee et al., 2009).

Of the 23 subjects in our study sample who completed the Detour-Reaching task, nine of the 15 marmosets (M1, M6, M8, M9, M10, M12, M13, M14, and M15) and 3 of the 8 squirrel monkeys (S3, S4, and S7) needed longer than five testing days to complete the five test sessions. These 12 subjects would have been excluded from (further) testing owing to motivational issues, had we stopped testing after five consecutive days as in opportunistic testing.

In our full testing, in which we gave the monkeys a total testing time period of 1 month to complete the Detour-Reaching task, all 12 individuals who needed more testing time than expected eventually also completed the task and no subject had to be excluded.

Task 2: Inhibition b – A-not-B

Measured cognitive ability. Inhibition of a learned routine, i.e. to avoid choosing the same location that had consistently been rewarded and switching to a second location instead that was now being rewarded. In order to successfully choose the correct cup, the subject had to inhibit a learned association between cup B and the food reward.

Test apparatus. Two opaque black plastic cups (A and B) were placed upside-down on a wooden sliding platform. The cups were attached to the board in a way that the experimenter could easily open and close them by flipping them back and forwards.

Familiarisation. The subjects learned (1) how to open the cups, (2) that only one cup at the time would contain a reward, and (3) that only one choice was possible in each trial.

Test. The test consisted of 6 trials conducted in a single session. During the first 5 trials, the experimenter placed a reward under cup B while the subject was watching, closed both cups simultaneously and quickly pushed the test apparatus within the subject's reach. The subject could then reach out to choose one of the two cups, by lifting it or attempting to do so. In the 6^{th} trial, the experimenter placed the cricket under cup A instead of B.

Cup A was only baited once a subject had correctly chosen cup B in 5 consecutive trials. If the subject did not correctly choose cup B in the first 5 consecutive trials, the experimenter repeated baiting cup B until the subject succeeded. If a subject did not make 5 consecutive correct choices within a single test session of 14 trials, it was tested again the following day. If a subject stopped participating during the session, it was also tested the following day.

Outcome variable. Correct choice in the last trial (cup A instead of B): yes/no.

Stop criterion in opportunistic testing. In opportunistic testing, testing with subjects who failed to choose cup B in the first 5 consecutive trials would have been stopped and those subjects would have been excluded from further analysis.

In our study, this would have resulted in losing 9 of the 22 subjects in the analysis who participated in the actual test. Eight of the 15 marmosets (M2, M4, M5, M8, M11, M13, M14 and M15) and one of the 7 squirrel monkeys (S8) did not achieve 5 correct B-choices in the initial 5 consecutive B-trials and therefore needed longer to succeed in 5 consecutive B-trials.

In our full testing approach in which we gave the monkeys one month (including the familiarisation phase) to complete the A-not-B task and as many B trials as needed, all 9 subjects who needed more than 5 B-trials to reach criterion eventually did so and completed the actual test. We only had to exclude one other squirrel monkey (S7) because he still refused to complete the familiarisation trials once all other subjects had completed the full task.

Task 3: Quantity Discrimination

Measured cognitive ability. Understanding of quantities by distinguishing between pairs of different quantities, i.e., two amounts of edible items.

Test apparatus. Schmitt & Fischer (2011) had reported that two monkey species performed best when the two quantities to choose from consisted of (1) inedible items (tokens), or (2) edible items that differed from the reward items in food type. In our study, we therefore used monkey chow pellets as stimulus items and crickets as a reward for a correct choice. Pellets are interesting enough to attract the marmosets' attention because they are edible but not desirable enough to distract them from the task at hand. However, in contrast to other studies (Schmitt & Fischer, 2011; Banerjee et al., 2009), the subjects in our study were not allowed to consume as many items as they had chosen, regardless of whether their choice was correct. Rather, the marmosets were rewarded with a single cricket if they correctly chose the larger pellet quantity in a given test trial and received no reward for an incorrect choice.

Pre-test. The subjects were presented with a single and obvious numerical contrast (1 vs. 6 pellets), and the location of the larger quantity was counterbalanced and pseudo-randomised for all contrasts. This ensured all subjects understood they had to choose the larger of two pellet quantities in order to obtain the cricket. Before proceeding to the test phase, they had to reach the criterion of $\geq 80\%$ correct trials within a single pre-test session of 10 trials.

Test. The test consisted of 3 sessions of 10 trials each. In each session, the subjects were presented with 10 different numerical contrasts (all possible quantity combinations of the 1-5 pellets). As in the pre-test phase, the location of the larger quantity was counterbalanced and pseudo-randomised for all contrasts.

Outcome variable. Per cent correct trials (choice of the larger amount) across all 3 test sessions.

Stop criterion in opportunistic testing. Under an opportunistic testing regime, testing would have been stopped if a subject did not complete the 30 trials within 3 sessions that were conducted on 3 consecutive days and this subject would have been excluded from the analysis.

Of the 15 marmosets who had entered the pre-test, 12 subjects participated in the actual Quantity Discrimination test. Nine subjects (7 females and 2 males) completed all 3 test sessions within the expected time frame of 3 testing days (one male subject even within 2 days). The remaining 3 male subjects needed up to twice as long to complete the task (M12 and M14 took 4 testing days, M6 took 6 days) and would have been excluded in opportunistic testing.

Under our full testing regime, in which we gave the monkeys one month (including the pre-test) to complete the task, the 3 above mentioned male individuals also completed the task. The only individuals we had to exclude were 3 other males (M1, M13, and M15) who were still not motivated to fully participate in the pre-test after a total testing period of one month and once their conspecifics had completed the whole task.

Task 4: Reversal Learning

Measured cognitive ability. Cognitive flexibility, i.e., reversing a previously learned association to a new one if the previously learned association is no longer rewarded.

Test apparatus. The apparatus consisted of a wooden board that contained two food wells and two rectangular wooden plates with two distinct black and white patterns printed on top.

Familiarisation phase. Each subject was familiarized with the basic setup and procedure in order to ensure it understood that it needed to uncover the baited well to obtain the reward by pushing away the correct plate. In this phase, both wooden plates looked identical (no patterns). In each trial, the experimenter placed a cricket in full sight of the subject in one of the two food wells and covered both wells with the plates. Pushing away the plate from the baited well allowed the subject to retrieve the cricket. To pass to the test phase, the subjects had to reach the criterion of 10/12 correct trials.

Test Phase 1 – Unhidden baiting – Learning of an association. The subjects were allowed to watch the baiting process to learn the association between one of the two patterns and the food reward. As in the familiarisation phase, the now patterned plates were initially each positioned behind one of the two food wells. The experimenter then placed a cricket in one food well and simultaneously covered both wells with the patterned plates. The reward was always under the same pattern and only choosing the correct one of the two patterns was rewarded. The criterion to move on to the next phase was again 10/12 correct trials.

Test Phase 2 – Strengthening of the learned association. The experimenter baited the food well with the same pattern as in phase 1 but now behind the occluder so that the subject could not observe the baiting process. As in phase 1, she simultaneously placed both patterned plates on top of the two food wells so that the subject could not make use of any auditory cues.

She then removed the occluder and the subject could make its choice. The criterion to move on to the next phase was again 10/12 correct trials.

Test Phase 3 – Reversal of the learned association. The experimenter followed the same procedure as in phase 2 with the crucial difference that she now always placed the food reward under the opposite pattern (i.e., the one that had previously not been rewarded). The criterion was again 10/12 correct trials.

Outcome variables. Number of sessions to reach criterion (80% correct trials within a single session) in each of the 3 test phases: (1) learning the association between one pattern and the food reward, (2) strengthening of the learned association and (3) reversal of the learned association. In addition, we used a cognitive flexibility measure, the Transfer Index (*TI*, Rumbaugh, 1997), which puts a subject's performance in the first session of phase 3 (reversal) in relation to its performance in the last session of phase 2.

Stop criterion in opportunistic testing. Had we used an opportunistic testing regime, subjects who did not complete each test phase in a maximum of 7 sessions would have been excluded from further testing. Since none of the marmosets in our study completed phase 3 within 7 sessions, this approach would have resulted in eventually losing the whole study sample.

In our full testing approach, we did not use such a strict criterion but we discontinued testing with subjects who (1) needed more than 19 test sessions of 12 trials in the last of the 3 test phases (reversal) or (2) failed to complete phase 2 or 3 owing to motivational issues over a total testing period of 5 months. Of the 15 marmosets who participated in the Reversal Learning task, only one male subject (M13) lost motivation very early on and failed to complete phase 1. Of the 14 remaining subjects, two males (M14 and M15), did not complete phase 2, despite having reached criterion very quickly in phase 1. Finally, of the 12 subjects who had completed phase 2 and entered phase 3, five subjects could not be tested until completion of the reversal: three males (M1, M6 and M8) who completely refused to participate owing to motivational issues and two females (M2 and M11) who developed a strong side bias in phase 3 which they did not overcome. However, since all 12 subjects had completed the first session of phase 3, the *TI* could be determined.

Task 5: Memory 1 – traditional task format with two choice options

Measured cognitive ability. Remembering in which of two locations a food reward has been hidden over increasing time delay periods.

Memory task 1a). In the traditional two-choice Memory task version 1a), the test apparatus and procedure largely corresponded to the Hidden Reward Retrieval test in Banerjee et al., 2009. Two white plastic containers were attached to a wooden board that could be slid into the subject's reach. Both containers were filled with dark bark mulch and a small piece of a yellow locust was used as a reward in each trial.

Memory task 1b). In the two-choice Memory task version 1b), the test apparatus consisted of two black containers with loosely placed lids that were attached to the middle one of three wooden platforms held together by a wooden frame (for details see Schubiger et al., 2016).

Pre-test. A no-delay condition was used as a pre-test. In this pre-test, the subjects could witness the baiting and choose immediately afterwards, without delay. Pre-test criterion was reaching ≥ 80% correct choices in a single 10-trial session. This ensured the subjects understood the task at hand before being exposed to time delay.

Test. At the beginning of each trial, both containers were covered with a rectangular piece of bark mull in Memory task 1a) or the containers were closed in Memory task 1b). Once the subject was attentive, the experimenter did the following: In Memory task 1a, she removed the cover of one of the containers, placed a piece of a locust in the container and put the cover back in place so that the bright yellow locust was no longer visible and both containers looked the same. In Memory task 1b, the experimenter lifted the lid of one of the two cups, placed a mealworm in the cup and closed it so that the mealworm was no longer visible to the subject. During the time delay period (5-30 seconds in Memory task 1a and 5-20 seconds in Memory task 1b), the test apparatus remained out of the subject's reach. As soon as the time delay had expired, the experimenter pushed the test apparatus into the subject's reach allowing it to choose one of the two containers. The subjects were tested with one session of 10 (Memory 1a) or 12 (Memory 1b) trials per day. A correct choice consisted of uncovering (Memory 1a) or opening (Memory 1b) the baited container.

Outcome variable: Number of correct trials (per cent correct choices) across all delay conditions in Memory 1a (6 delays, 60 trials) and Memory 1b (4 delays, 48 trials), as well as in both task versions combined (4 delays, 40 or 48 trials, all subjects from a and b).

Stop criterion in opportunistic testing.

Memory 1a). In opportunistic testing, subjects would have been excluded if they did not complete all 6 test sessions in the expected 6 testing days (Hidden Reward Retrieval, Banerjee et al., 2009). Since none of the subjects in marmoset sample 1 who participated in this two-choice memory task fulfilled this criterion, they would all have been excluded from further testing and no statistical analysis would have been possible.

However, the majority of subjects completed the task within 12 testing days which we set as expected testing time instead.

In our full testing approach, we gave each marmoset as many test sessions as it needed to complete the pre-test and the 6 test sessions during a 5-month testing period. Of the subjects who needed more than 12 testing days, five (M1, M3, M6, M8, and M13) eventually completed Memory task 1a. However, three other marmosets (M14, M15, and M16) had not completed the whole task once the 5-month testing period expired and had to be excluded from statistical analysis.

Memory 1 b). In opportunistic testing, subjects would have been excluded if they failed to complete the 4 test sessions in the expected 4 testing days. This would have resulted in excluding 4 of the 8 marmosets (M18, M19, M20, and M24) and one of the 8 squirrel monkeys (S3) who participated in Memory task 1b).

We gave each subject a 5-month testing period (including the pre-test) to complete the memory task. Fifteen of the 19 monkeys (8 of the 11 marmosets of study sample 2 and 7 of the 8 squirrel monkeys) who had entered Memory task 1b eventually completed the full task within the given testing period. The remaining 3 marmosets (M22, M26, and M27) and one squirrel monkey (S6) had lost motivation early in the pre-test or after a very small number of test trials and the experimenter could not regain their motivation to fully participate. Using our full testing approach, we thus only had to exclude these 4 subjects from statistical analysis who were not sufficiently motivated to participate in the actual test.

Since the two versions of Memory test 1 slightly differed in terms of used methodology which could potentially have affected the results, we first analysed and reported them separately. In a third analysis, we then also combined the subjects of marmoset sample 1 who had completed the first 4 test sessions of Memory task 1a) and the subjects of marmoset sample 2 and the squirrel monkeys who had completed Memory task 1 b).

Using our full testing approach in which we had given each subject sufficient time to complete the (first) four delay conditions of Memory task 1, we obtained performance scores for one more male marmoset from sample 1 (M14) who had dropped out in Memory task 1a because they had not completed all 6 delay conditions. Therefore, we only had to exclude two subjects of study sample 1 (M15 and M16) who had not completed all trials of the 4 test sessions after expiration of the 5 months testing period.

Three other marmosets (M22, M26, and M27) of study sample 2 and one squirrel monkey (S6) also had to be excluded because they had dropped out in the pre-test phase or after only a few test trials owing to motivational issues.

Task 6: Memory test 2 – Optimised test format with nine choice options

Measured cognitive ability. Remembering in which of nine locations a food reward has been hidden over increasing time delay periods.

Apparatus and procedure. This memory task differed from task 5 in that nine rather than two options were available to choose from. The apparatus and procedure were identical to the ones used in Memory task 1b) with the only difference being that 9 rather than only 2 cups were used. The method has been described in detail in an earlier publication (Schubiger et al., 2016).

Outcome variable. Per cent correct choices across all delay conditions (4 delays, 48 trials).

Stop criterion in opportunistic testing. In opportunistic testing, subjects would have been excluded if they had not completed the 4 delay conditions in 4 sessions conducted on 4 consecutive days.

In our full testing approach, we continued testing for a total testing period of 5 months in order to allow every subject to complete the test. Seven of initially 11 marmosets and 7 of initially 8 squirrel monkeys completed the whole Memory task 2 whereas one marmoset (M24) only completed the first test session (the 5-second delay condition). Three marmosets (M22, M26, and M27) and one squirrel monkey (S6) eventually had to be excluded because they had either not completed the pre-test or only a very small number of test trials before they refused to participate at all.

4.4 Data analysis

For each of the six cognitive tasks, we analysed whether individuals who finished the task in the expected time (i.e., would have been included in opportunistic testing) differed in performance from individuals who needed longer (i.e., would have been excluded in opportunistic testing but were included in our full testing). We ran Generalised Linear Mixed Models (GLMMs), with the exception of the A-not-B task (task 2) for which we computed Fischer exact tests owing to the small number of trials in this task. The outcome variables for cognitive performance were the response variables, and subject was included as a random factor in all models. Testing time needed to complete a given test was included as a binary fixed factor: expected amount of time vs. longer. Individual test session, species (where applicable) and all two-way interactions were also included as fixed factors in the models. We calculated all models with biologically meaningful factor combinations and identified the best model using the Akaike criterion corrected for small sample sizes (AICc, Hurvich & Tsai, 1998).

4.5 Results

4.5.1 Detour-Reaching

Of the 23 individuals who completed the Detour-Reaching task, 11 individuals completed the task in time (≤ 5 days; 6 marmosets, 5 squirrel monkeys) whereas 12 individuals took longer than expected (> 5 days; 9 marmosets, 3 squirrel monkeys). The best model to explain performance in the Detour-Reaching task included the fixed factors test session ($F(4, 84) = 8.48$, $p = .000$), species ($F(1, 21) = 1.26$; $p = .274$), and their interaction (session*species: $F(4, 84) = 5.51$, $p = .001$). Thus, individuals who finished the task in the expected time did not differ in performance from individuals who took longer to complete it (see Figure 4.1). The interaction session*species indicates that while the marmosets improved their performance over the course of the five sessions and therefore learned to inhibit to directly reach for the food, this was not the case for the squirrel monkeys (see Figure S4.2).

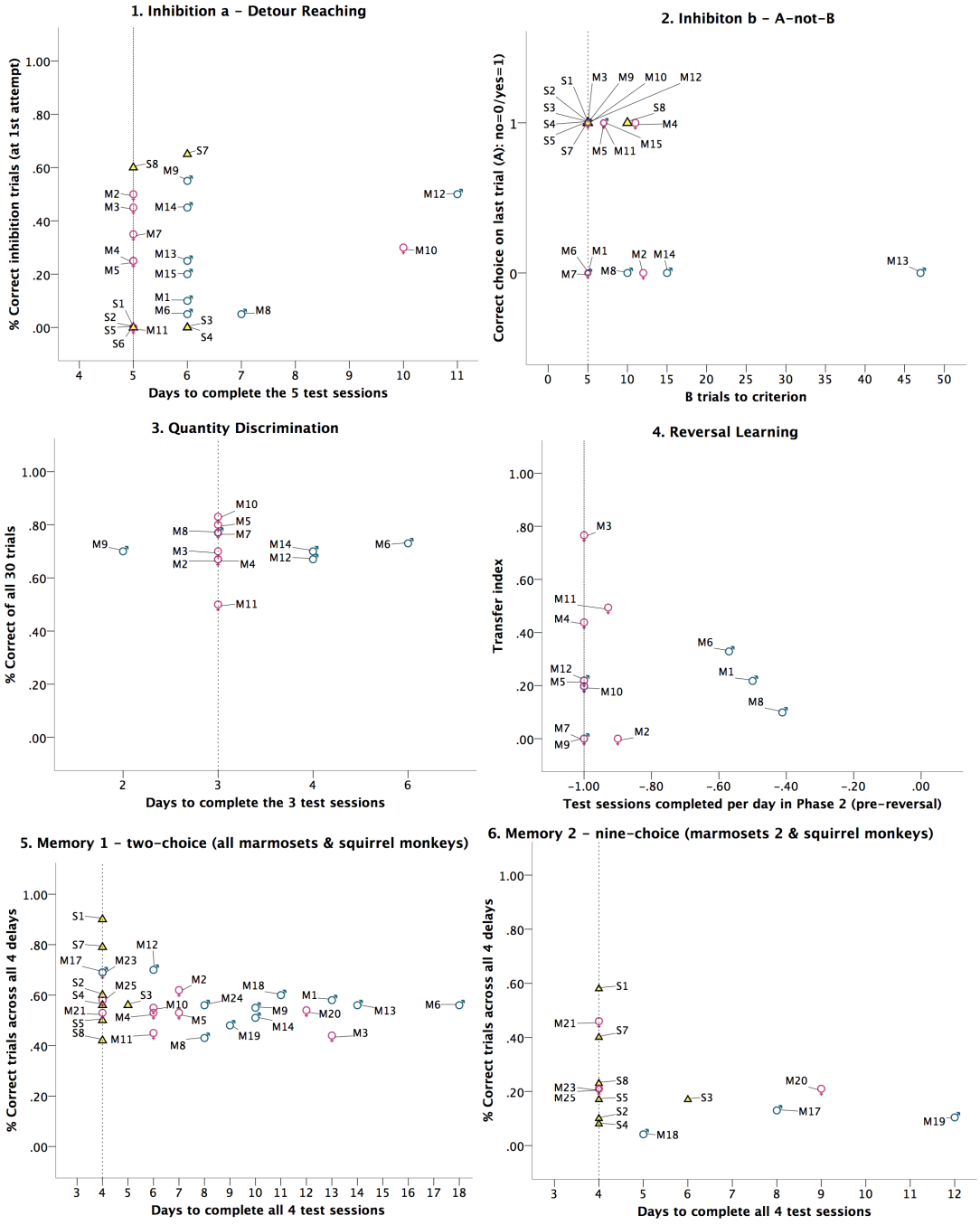


Figure 4.1 The subjects' performance in the 6 cognitive tasks depending on how much testing time they needed to complete each task. The dotted line marks the expected amount of time after which testing would have been discontinued and subjects who needed longer would have been excluded had we used the classical opportunistic approach. Marmosets are represented by ID numbers M1-M27 and gender symbols, squirrel monkeys by ID numbers S1-28 and yellow triangles.

4.5.2 A-not-B

Of the 22 individuals who completed the A-not-B task, 13 individuals (7 marmosets, 6 squirrel monkeys) completed the test trials in time (five B-trials to criterion). Nine individuals (8 marmosets, 1 squirrel monkey) took longer than the expected amount of testing time (> five B-trials) to reach criterion (5 consecutive correct B-trials) before receiving the final (A) trial. Overall, 8 of the 15 marmosets and all 7 squirrel monkeys who completed the A-not-B task correctly chose cup A in the last trial and therefore passed the task whereas the remaining 8 marmosets failed it. Since a GLMM was not applicable owing to the relatively small number of test trials, two Fisher exact tests were conducted instead. The first Fisher exact test determined whether testing time (the number of B trials required to correctly choose cup B in five consecutive trials) affected the subjects' performance (whether they correctly chose cup A in the last trial). The second test determined whether the species difference in performance was significant. Although there was a non-significant trend that subjects who needed more time to complete the A-not-B task were more likely to fail it, testing time had no significant effect on the monkeys' performance (Fisher's exact test; $p = 0.178$). Likewise, there was a non-significant trend for squirrel monkeys to outperform marmosets (Fisher's exact test; $p = .121$) (see Figure 4.2).

4.5.3 Quantity Discrimination

Of the 12 marmosets who completed the Quantity Discrimination task, nine individuals did so in time (≤ 3 days; 7 females, 2 males) whereas 3 individuals took slightly longer than expected (> 3 days; 3 males). Overall, the 12 subjects who completed the Quantity Discrimination test performed well above chance (*Mean score* = 71% correct choices, *SD* = 8%; $t(11) = 8.53$, $p = .000$). The best model to predict cognitive performance was the one that only included one fixed effect, testing time, which did not affect the marmosets' performance ($F(1, 10) = 0.42$; $p = .534$) (see Figure 4.3).

4.5.4 Reversal Learning

For the Reversal Learning task, we first report how long (number of test sessions) the subjects needed to reach the learning criterion (≥ 80% correct trials within a single session) in each of the 3 test phases. We also report whether the number of days the subjects needed on

average to complete a 12-trial test session affected their performance (per cent correct trials in the last session). Finally, we answer the central question: whether the amount of testing time (in this task measured as the number of completed trials per testing day) affected the subjects' cognitive flexibility (as measured by the *TI*).

Phase 1 - Learning of an association

Of the 14 marmosets who completed phase 1, all individuals did so in time whereas no subject took longer than expected. In phase 1, the 14 marmosets reached criterion (≥ 80% correct trials in a single session) within an average of 2 test sessions (M = 2.07, SD = 1.60), with the minimum being 1 session and the maximum 6 sessions (see Figure S4.3).

Phase 2 - Strengthening of the learned association

Of the 12 marmosets (7 females and 5 males) who completed phase 2 (strengthen the learned association), seven (4 females and 3 males) did so in time whereas 5 (3 females and 2 males) needed longer than expected. The number of 12-trial sessions needed to complete phase 2 ranged from 4 to 13 sessions, with the average being 7 sessions (M = 6.75, SD = 3.11) (see Figure S4.3).

Phase 3 Reversal of the learned association

Of the 7 marmosets (4 females and 3 males) who completed phase 3 (full reversal of the initially learned association), only 1 female subject (Vesta) did so in time whereas 6 subjects (4 females and 2 males) needed longer. The marmosets completed phase 3 in 7 to 19 sessions, with the average being 13 sessions (M = 12.57, SD = 4.96) (see Figure S4.3).

Transfer Index (*TI*)

All of the 12 marmosets who had entered phase 3, completed at least the first test session of the reversal. To compare cognitive flexibility between individuals, the *TI* was calculated for each of these subjects. In order to determine whether testing time, i.e., the number of days a subject needed to complete a session in phase 2 (expected = 1 session of 12 trials per day; longer = < 1 session of 12 trials per day) or a subject's sex had an effect on its cognitive flexibility (as measured by the *TI*), we ran GLMMs with the fixed effects time, sex, and the interaction of these two factors. The two models that only included testing time and sex, respectively, only differed marginally from each other (ΔAICc < 1). Neither time (F(1, 10) = 0.62, p = .808) nor sex (F(*1*, 10) = 0.91, p = .362) significantly affected a subject's *TI* (see Figure 4.4).

4.5.5 Memory 1 – Traditional two-choice task

Version a) Marmoset sample 1

Of the 12 marmosets (6 females and 6 males) who completed the 2-choice Memory task 1 a (study sample 1), 6 subjects did so in our set time (≤ 12 days) whereas 6 subjects took longer (> 12 days). The best model to predict whether time needed to complete the task had an effect on the monkeys' performance only included the fixed effect time which did not have a significant effect on performance (per cent correct trials): $F(1, 10) = 0.71$; $p = .420$. The two models that only differed marginally from the best model (ΔAICc < 1) each also included only one fixed effect, delay and sex respectively. Delay condition had a significant effect on a subject's performance ($F(1, 5) = 3.44$; $p = .009$) whereas a subject's sex did not ($F(1, 10) = 0.17$; $p = .690$) (see Figure S4.4a).

Version b) Marmoset sample 2 & squirrel monkeys

Of the 15 monkeys (8 marmosets and 7 squirrel monkeys) in study sample 2 who completed the two-choice memory task, eleven individuals (5 marmosets, 6 squirrel monkeys) did so in time (≤ 4 days) whereas the other 5 individuals (4 marmosets, 1 squirrel monkey) needed longer (> 4 days).

The best model only included the fixed factor time which did not have a significant effect on the monkeys' performance ($F(1, 18.66) = 0.38$; $p = .544$) (see Figure S4.4b).

Extended study sample (all subjects from both task versions a and b)

When combining the (first) four delay conditions of those marmosets (sample 1 and 2) and squirrel monkeys who had completed all essential trials of the two-choice memory task, we obtained performance scores for two more male marmosets from sample 1 (Kapi and Kantor) who had dropped out in Memory task 1a because they had not completed all 6 delay conditions. Of the 28 monkeys (7 squirrel monkeys and 21 marmosets) who completed the four delay conditions, ten (4 marmosets and 6 squirrel monkeys) did so within the expected time (4 days) whereas 18 (1 squirrel monkey, 17 marmosets) took longer (> 4 days). The best model only included the fixed factor delay which had a highly significant effect on the monkey's performance ($F(3, 81) = 5.33$, $p = .002$). Although the next best model also included time, required testing time did (as in all other models) not significantly affect performance ($F(1, 26) = 3.00$, $p = .095$) (see Figure 4.5).

4.5.6 Memory 2 – Optimised nine-choice memory task (marmosets 2 & squirrel monkeys)

Of the 14 monkeys (7 marmosets of study sample 2 and 7 squirrel monkeys) who completed the nine-choice memory task, nine individuals (3 marmosets, 6 squirrel monkeys) did so in time (4 days) whereas the other 5 individuals (4 marmosets, 1 squirrel monkey) needed longer (> 4 days). Time to complete the 9-choice memory task (4 days versus > 4 days) did not affect the subjects' performance in the 9-choice memory task (no significant effect in any of the models). The best model only included the fixed effect delay which did affect performance (i.e. the proportion of correct choices): $F(1, 3) = 5.48$; $p = .003$, indicating that memory performance decreased with increasing retention delay. While species had no effect on overall performance in any of the models, there was a significant interaction of delay and species in the third-best model (see Figure 4.6).

4.6 Discussion

We tested marmosets and squirrel monkeys with six cognitive tasks of a modified version of a cognitive test battery that had been developed for New World monkeys (Banerjee et al., 2009). These were Detour Reaching, A-not-B, Quantity Discrimination, Reversal Learning and two memory tasks. We used a full testing approach and tested all available individuals. This approach is in contrast to the commonly used opportunistic testing approach in which subjects are excluded if they fail to complete a cognitive task (or pre-test) within a predefined time frame. Our full testing approach thus allowed each subject ample opportunity to complete each task at its own pace. The main question was whether the amount of testing time a subject required to complete a task affected its cognitive performance in that task. Our results show that this was not the case and that in each task the performance of subjects who needed longer than expected to complete the task did not differ from those who completed the task in time.

It is important to stress that even in our full testing approach, we lost subjects from the sample. This was because the monkeys were never forced to enter the testing enclosures but had to be at least sufficiently motivated to approach the testing area voluntarily to be included. While we had made every effort to allow each individual to complete each task at its own pace, it was not justified to continue testing for a completely unlimited time period. Therefore, we cannot rule out the possibility that the few subjects who dropped out in some tasks despite

being given ample time to complete them, differ in cognitive ability from subjects who completed the tasks.

Despite this possible exception, our findings suggest that opportunistic testing in primates (i.e. only testing subjects who readily participate) does not bias the results of cognitive tasks in several physical domains, at least in marmosets and squirrel monkeys. To what extent they generalise to tasks from the social domain or to paradigms that do not require subjects to choose between a set of presented options remains to be established.

Species differences in cognitive performance were minor and mostly non-significant, even though we cannot exclude that this may be owing to the relatively low number of squirrel monkeys in our sample. The only significant effect was that in the detour reaching task, marmosets increased their performance over time, whereas squirrel monkeys did not, which led to increased inhibition performance in marmosets in later trials (see Figure S4.1). One possible explanation is that the marmosets are better at inhibiting a direct reach for food because, as cooperative breeders, they frequently share food with immature group members including food offering and calling others to a food source (Brown, Almond, & Bergen, 2004). Food sharing, and in particular proactive food offering, arguably requires the ability to inhibit the immediate impulse to take and consume the food. To further test this possibility, future tasks could compare whether marmosets who are more inclined to share food indeed show stronger detour reaching performance. Intriguingly, in the second inhibition task, the A-not-B task, marmosets did not outperform the squirrel monkeys but there was rather a trend in the opposite direction. It thus appears that the two inhibition tasks measure different aspects of inhibitory control. This possibility is consistent with Mac Lean et al. (2014) who tested a large number of primate (and other) species with the same two inhibition tasks and likewise did not find a particularly strong correlation between the two.

Our finding that the selection bias that results from opportunistic testing does not affect cognitive performance in marmosets and squirrel monkeys in several physical cognition tasks adds to the list of non-cognitive factors that have been studied so far regarding their potential to affect cognitive performance (Table 4.1). While required testing time did not affect cognitive performance in our study, another testing-related factor, task format, did affect cognitive performance of non-human primates (Schmitt & Fischer, 2011; Carducci et al., 2018) and it sometimes did so differently in different species (Schmitt & Fischer, 2011; Gazes et al., 2017).

Although the effects of subject-related factors (such as individual differences in personality traits or housing conditions) seem less ambiguous, the overall pattern remains inconclusive.

To sum up, we showed that at least for two New World monkey species and the given set of cognitive tasks, opportunistic testing does not bias the test results. If our findings generalise to other non-human primate species and cognitive tasks, then maximising sample sizes by only testing consistently motivated subjects will be a valid alternative whenever a sufficiently large number of subjects is available but full testing is not feasible owing to time constraints.

Future studies should extend the list of non-cognitive factors that affect cognitive performance, so that the validity and reliability of existing cognitive tests can be further improved and newly developed tests will be less prone to performance biases. Once this task is achieved, cognitive test batteries can then be used more reliably to compare the cognitive abilities of non-human primate individuals and species.

4.7 Ethics Statement

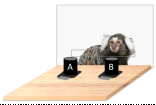
This study was performed in accordance with the Swiss legislation and licensed by the Veterinary Office of the Canton of Zurich (Licence number 183/13, 24826, degree of severity: 0, i.e., no harm). Thus before, during and after this study, the monkeys were never constrained or subjected to any pain, suffering or injury and their general state of health was not impaired. All cognitive tasks were conducted non-invasively between the monkeys' regular feeding times. The monkeys could freely enter and leave the test enclosure without being handled by humans at any time and were never isolated from their social groups. After the completion of this study, the monkeys continued living at the Primate Station, eventually participating in other non-invasive studies.

4.8 Acknowledgements

We thank our research interns, Piero Amodio (squirrel monkeys) and Chantal von Siebenthal (marmosets), who helped with habituation, and our animal husbandry team at the Primate Station, Heinz Galli, Claudia Rudolf von Rohr, Thomas Bischof, and Patricia Rivera, who took care of the monkeys during the course of this study. We are also grateful to Heinz Galli for help with building the experimental apparatuses. Furthermore, we thank Carel van Schaik and Nick Mulcahy for constructive discussion, and Erik Willems for statistical advice. Finally, we thank the three anonymous reviewers for their helpful comments and suggestions.

4.9 Supporting Online Information

Table S4.1 Materials & measurements for the tasks of the cognitive test battery.

Task	Item	Measurements (in cm)	
		Marmosets	Squirrel monkeys
1.Detour-Reaching 	Window front (1)	18.0 x 2.5	
	Wooden board with an indentation (1cm deep) for Perspex panel	34.0 x 37.0	42.0 x 46.0
	Transparent Perspex panel (0.5 cm in thickness); quadratic surface exposed	8.0 x 9.0	13.5 x 14.5
	Reward: mealworm (or cricket)		
2. A-not-B 	Window front (1)	18.0 x 2.5	
	Wooden board	34.0 x 37.0	42.0 x 46.0
	2 black plastic cups (each with a silver insulating strip on its base)	Ø 2.6, h. 3.8	Ø4.5, h: 5.2
	Reward: mealworm (or cricket)		
3. Quantity Discrimination 	Window front (2)	18.0 x 2.5	
	Wooden board with 4 wooden sliding rails for sliding platforms	41.0 x 40.0 2.0 x 33.0	
	2 transparent plastic sliding platforms with white plastic handles on their front ends	28.0 x 7.0 1.0 x 7.0	
	2 small petri dishes with white floors (plastic foil) and transparent lids	Ø 5.2, h: 1.2 Ø 5.6, h: 0.7	
	2 white flat cotton pads to keep the pellets in place	Ø 5.0	
	Experimental stimuli: monkey chow pellets	Ø 0.5, l: 1.5	
	Rewards: crickets		
4. Reversal Learning 	Window front (2)		
	1 wooden sliding board with 2 round food wells at 2 cm distance from board's front	33.5 x 32.5 Ø 5.1, h: 1.6	
	2 wooden plates, each with pattern (=paper covered with plastic foil): 1) white triangle on black background 2) alternating black and white "waves"	7.5 x 6.5	
	Rewards: crickets		
5. Memory 1a 	Window front (2)		
	1 wooden sliding board with 2 round food wells at 2 cm distance from board's front	33.5 x 32.5 Ø 5.1, h: 1.6	
	2 white round plastic containers (11 cm apart) filled with ramial chipped bark mulch	Ø 5.1, h: 3.1	
	2 cover plates (= larger rectangular bark mulch pieces of equal size)		
	Rewards: yellow desert locusts		
5. Memory 1b 	Wooden sliding board	45.0 x 30.0	95.0 x 50.0
	Wooden frame with 3 platforms	w: 40 x h: 37.5	W: 80.0 x h:75.0
	2 black plastic cups with grey lids	Ø 3.1, h: 1.1	Ø 3.1, h: 2.3
	2 Velcro tape strips		
	Rewards: crickets or small pieces of a cooked apple (M); mealworms or small pieces of cashew nut (S)		
6. Memory 2 	Wooden sliding board	45.0 x 30.0	80.0 x 75.0
	Wooden frame with 3 platforms	w: 40 x h: 37.5	w: 80.0 x h: 75.0
	9 black plastic cups with grey lids	Ø 3.1, h: 1.1	Ø 3.1, h: 2.3
	9 Velcro tape strips		
	Rewards: crickets or small pieces of a cooked apple (M); mealworms or small pieces of cashew nut (S)		
	For a detailed description of the memory tasks see Schubiger et al. (2016)		

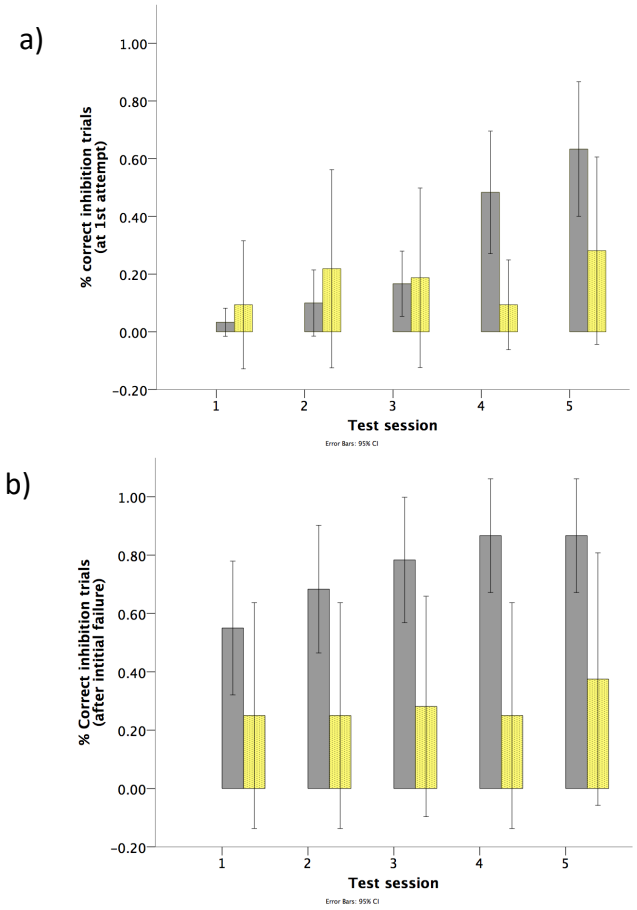


Figure S4.1 Performance in the 5 test sessions of the Detour-Reaching task. Successful detour-reaching in the inhibition trials at (a) first attempt and (b) after initial failure to reach around the transparent barrier. Marmosets (grey solid bars)/squirrel monkeys (yellow dotted bars).

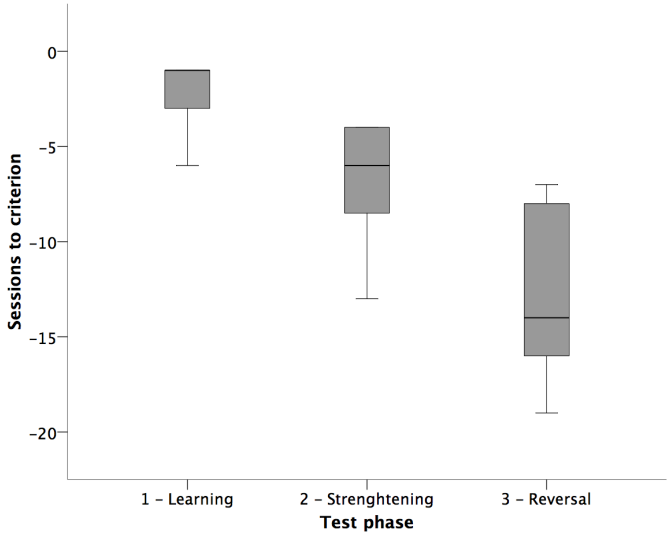


Figure S4.2 Number of test sessions to criterion in the 3 phases of the Reversal Learning task. Fewer (12-trial) sessions indicate better performance, hence the negative prefix on the y-axis. Phase 1 - learning of the initial association between a pattern and a reward. Phase 2 - strengthening of the learned association. Phase 3 - reversal of the learned association.

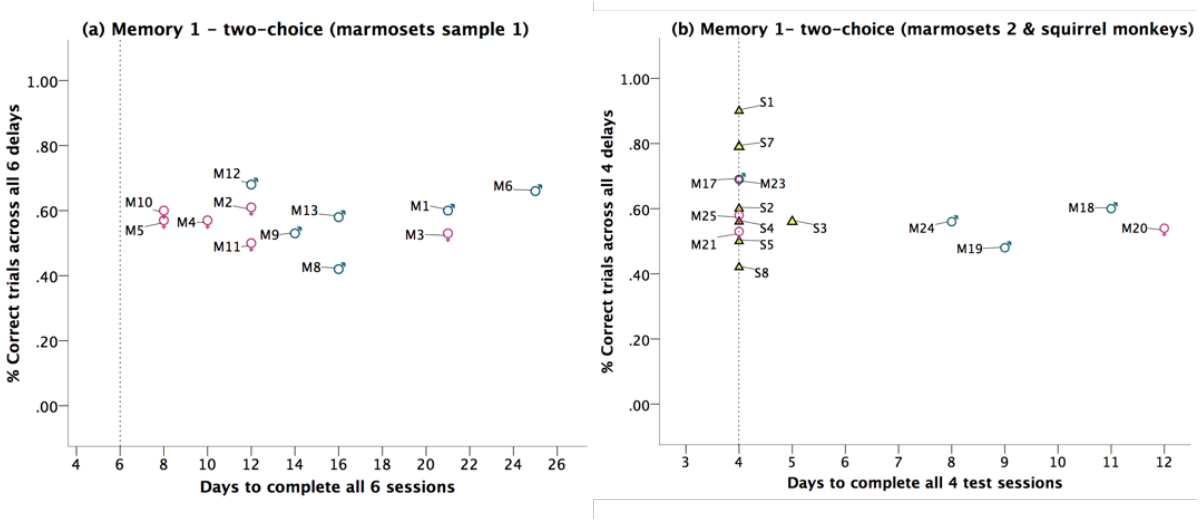


Figure S4.3 The subjects' performance in the two versions of Memory task 1 depending on how much testing time they needed to complete the task. The dotted line represents the expected amount of time (number of testing days) after which testing would have been discontinued and subjects who needed longer would have been excluded had we used the classical opportunistic approach. (a) The marmosets of study sample 1 (represented by ID numbers M1-M16 and gender symbols) were tested with version a of Memory 1 (adapted from Banerjee et al., 2009), (b) the marmosets of study sample 2 (ID numbers M17-M27 and gender symbols) and the squirrel monkeys (represented by yellow triangles) were tested with version b (Schubiger et al., 2016). The two versions of Memory 1 slightly differed in details of experimental set-up (apparatus) and design (number of delay conditions; also see Table 4.3 and S1 Table S4.1).

Chapter 5

General Discussion

5.1 Summary of findings

In this thesis, I aimed at identifying if current tests of primate cognition allow for fair comparisons of cognitive abilities within and between non-human primate species and how they can be improved, if necessary, in order to help us gaining insight into how the human mind might have evolved. Until recently, such studies with non-human primates have mainly focused on between-species rather than within-species comparisons of cognitive abilities without questioning the validity of the used cognitive tasks. However, in order to be able to compare the cognitive abilities of different species in a fair manner, it seems important to first attempt to better understand in how far individuals of a single species vary. In my thesis, I used this modern approach by primarily focusing on a single New World primate species, common marmosets (*Callithrix jacchus*). In addition, I conducted first comparisons to common squirrel monkeys (*Saimiri sciureus*), the closest evolutionary relatives of marmosets, apart from other callitrichids. My main goal was to assess the internal validity of a set of commonly used cognitive tasks (i.e., whether they truly measure individual differences in cognition rather than non-cognitive factors). In order to do so, I conducted a series of three experimental studies that addressed if several non-cognitive factors (i.e., individual differences in emotional reactivity, task format, and time to complete a task) affected the monkeys' performance.

5.1.1 Individual differences in emotional reactivity affect motivation but not cognitive performance

In my first study, I investigated if individual differences in levels of emotional reactivity affect performance in cognitive tasks. In order to elicit variation in emotional reactivity, individual marmosets were tested with an object permanence test by the same human experimenters who were either familiar or unfamiliar to the marmosets. Unfamiliar experimenters were more likely than familiar experimenters to induce emotional arousal in the marmosets during testing because the marmosets were not habituated to unfamiliar experimenters. The results showed considerable individual variation in marmosets' spontaneous emotional reactivity. Male subjects, especially, tended to experience elevated emotional arousal which they behaviourally expressed in piloerection and repeated vigilance and mobbing vocalisations. They also tended to avoid approaching the test stimuli and to indicate their desire to leave the test situation. As a consequence of this, highly emotionally

reactive male subjects participated in fewer test trials and tended to be easily distracted during the test procedure. However, in those trials in which they did participate and were attentive to the task, their cognitive performance was not affected by emotional reactivity levels (Figure 5.1).

Figure 5.1 Findings of Study 1. A subject's emotional reactivity (arousal and avoidance behaviour) in the test situation affected its participation (number of completed test trials) but not its performance (number of correct trials) in the object permanence task.

My findings illustrate that differences in emotional reactivity have no major impact on subjects' cognitive performance. Importantly, however, this only applies if performance is controlled for whether the subjects were attentive during the stimulus presentation, and if clear stop criteria are applied that determine under which circumstances a test (session) has to be discontinued. For the marmosets, this was the case when a subject repeatedly refused to make a choice (in 4 consecutive test trials) or clearly avoided the test situation by repeatedly climbing to the test compartment's entrance/exit door and trying to open the door to leave. While such stop criteria prevent performance biases in cases of particularly high emotional arousal, giving subjects the possibility to discontinue a test session at any time is also essential for animal welfare reasons.

When testing marmosets and possibly other non-human primates, experimenters have to be aware that particularly male subjects will often be vigilant towards their environment and attend to visual and auditory stimuli other than the test stimuli. In such cases, the experimenter will have to regain the subject's attention and ensure it is watching the test procedure before making a response (e.g., a choice). In studies in which this is only possible to a certain degree, such as mine in which it was essential for comparability reasons that all experimenters continuously proceeded with their actions, it is advisable to later code from the video recordings in which trials the subject was truly attentive. This will allow researchers to correct the subjects' performance scores accordingly by only analysing trials in which the subjects were attentive in the test.

5.1.2 Some widely used memory tests are not valid

In my second study, I addressed construct validity, i.e., if cognitive tests measure the cognitive ability (i.e., the cognitive construct) they were designed to measure. Specifically, I explored whether one aspect of the task format (the number of test stimuli to choose from) affected the performance of marmosets and squirrel monkeys in a typical long-term memory test. This was achieved by testing both species with two versions of the memory task: a traditional two-choice and a novel nine-choice version. When comparing the performance of the two species on the same two tasks, the results suggested that increasing the number of choice options from two to nine improved their motivation to choose correctly. And this in turn had a positive effect on their performance: they performed better in the nine-choice memory test than the two-choice version (Figure 5.2). One likely explanation for this finding is that subjects may have used a random selection strategy in the two-choice task but not in the nine-choice task because there was more at stake in this version by choosing randomly. There was only a 11.1% chance that the subject could find the reward by choosing at random compared to a 50% chance in the two-choice task. The most important result was, however, that in the nine-choice test version, the monkeys' performance decreased as the long-term memory demands increased. The fact that their performance in the nine-, but not in the two-choice task mapped the Ebbinghaus forgetting curve suggests that only the novel nine-choice memory test truly measured long-term memory and was construct-valid whereas the two-choice test was not.

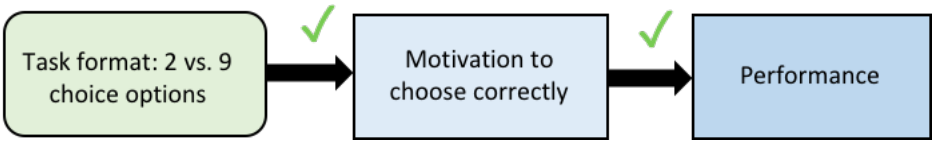


Figure 5.2 Findings of Study 2. The number of choice options affected motivation and performance. Subjects performed better in a memory test with nine than with two options and only in the nine-choice task their performance decreased with longer time delays during which the reward's location had to be remembered.

My findings that relatively small changes to the task format (i.e., increasing the number of available choice options) promoted the motivation of marmosets and squirrel monkeys to perform well in a memory test are an example of how the validity of cognitive tests can be improved.

5.1.3 Individual differences in time to complete a cognitive test do not necessarily affect performance

In my third study, I investigated the potential effect of individual differences in the amount of required testing time on individual differences in performance. In particular, I assessed whether the common opportunistic testing approach that excludes subjects who need longer than the scheduled amount of testing time to complete a task leads to a bias in the performance results. For this purpose, I tested marmosets and squirrel monkeys with a full testing approach that allowed individuals to complete the tasks of the cognitive test battery in more than the scheduled testing time/days. I then compared the performance of subjects who needed more than the scheduled testing time, i.e., who would typically be excluded from testing, with those subjects who completed the tasks in the scheduled time and would therefore not be excluded in opportunistic testing. The results revealed that the performance scores of the two groups of subjects did not differ in any of the cognitive tasks (Figure 5.3). Therefore, although opportunistic testing leads to a bias in selecting which subjects participate in a test based on their motivation, it does not necessarily lead to a bias in the results.

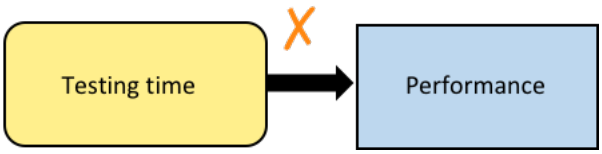


Figure 5.3 Findings of Study 3. Individual difference in required testing time did not affect the subjects' performance in any of the cognitive tasks of the test battery.

Based on my findings, opportunistic testing can be considered as an acceptable alternative to full testing when the latter is not feasible owing to external circumstances. This is an important finding because the central issue that researchers working with non-human primates often have to deal with is limited access to test facilities that restricts the time frame of an experimental study. Therefore, opportunistic testing can be a reasonable option in such situations, at least in marmosets and squirrel monkeys, as it may have no or little effect on the results and should produce similar findings compared to the more time-consuming full testing approach.

5.1.4 Overview of findings

As summarised in Figure 5.4, only one of the three non-cognitive factors, task format, seriously affected the cognitive performance of the two New World monkey species I tested. The other two factors, a subject's emotional reactivity and the amount of testing time required to complete a task did not.

In order to illustrate which potentially confounding factors can be best controlled by researchers, I had divided those non-cognitive factors into external ones (related to the test design) and internal ones (related to the subject's state). However, the two kinds of factors overlap to some degree. The external factor 'opportunistic testing', for instance, is a way in which the experimenter deals with an internal factor, the subject's required 'testing time'. How much time a subject requires to complete a cognitive task depends on its 'motivation to participate', which in turn is partly affected by its 'emotional reactivity' levels that also affect its attention in the cognitive task. When a subject is too emotionally aroused or ceases to attend and respond to the cognitive task, the experimenter has to stop a test session (and continue testing on another day) in order to avoid that the subject's behaviour might negatively affect cognitive performance. If some subjects repeatedly react in this way, cognitive testing can become so time-intensive that the experimenter eventually has to exclude the subject from further testing (opportunistic testing). Since experimenters have little influence on the subject's emotional reactivity levels, or its motivation to participate, apart from attempting to closely monitor and regain its attention, they can only ensure that the subject's performance is corrected accordingly. Methodological aspects, however, that might affect a subject's motivation and performance, can be better controlled via careful test design and by evaluating and optimising the task's measurement validity and reliability.

In this thesis, I conducted such an evaluation by assessing the measurement validity of the test battery's original memory task. My results suggested that two New World monkey species used a random-choice strategy in the two-choice memory task rather than remembering the location of the food reward during delay periods. Although this non-memory strategy does not allow the subject to locate the reward 100% of the time, it does give the subject a 50% chance of finding the reward on each trial rather than making the effort of remembering the location of the food reward during the delay periods. In the novel nine-choice version of the memory test, however, a subject's strategy to choose at random resulted in a very low success quote of just 11.1% in each test trial. It therefore paid off to correctly

remember the reward's location in the 9-choice version which was reflected in the monkeys' better performance and their increasing difficulty to remember the reward's location as the time delays got longer and the long-term memory demands increased. Therefore, this new test version enhanced the memory test's validity substantially. Importantly, a task format with more than the usual two or three options will likely enhance the validity of other cognitive tests with inherently low working memory demands.

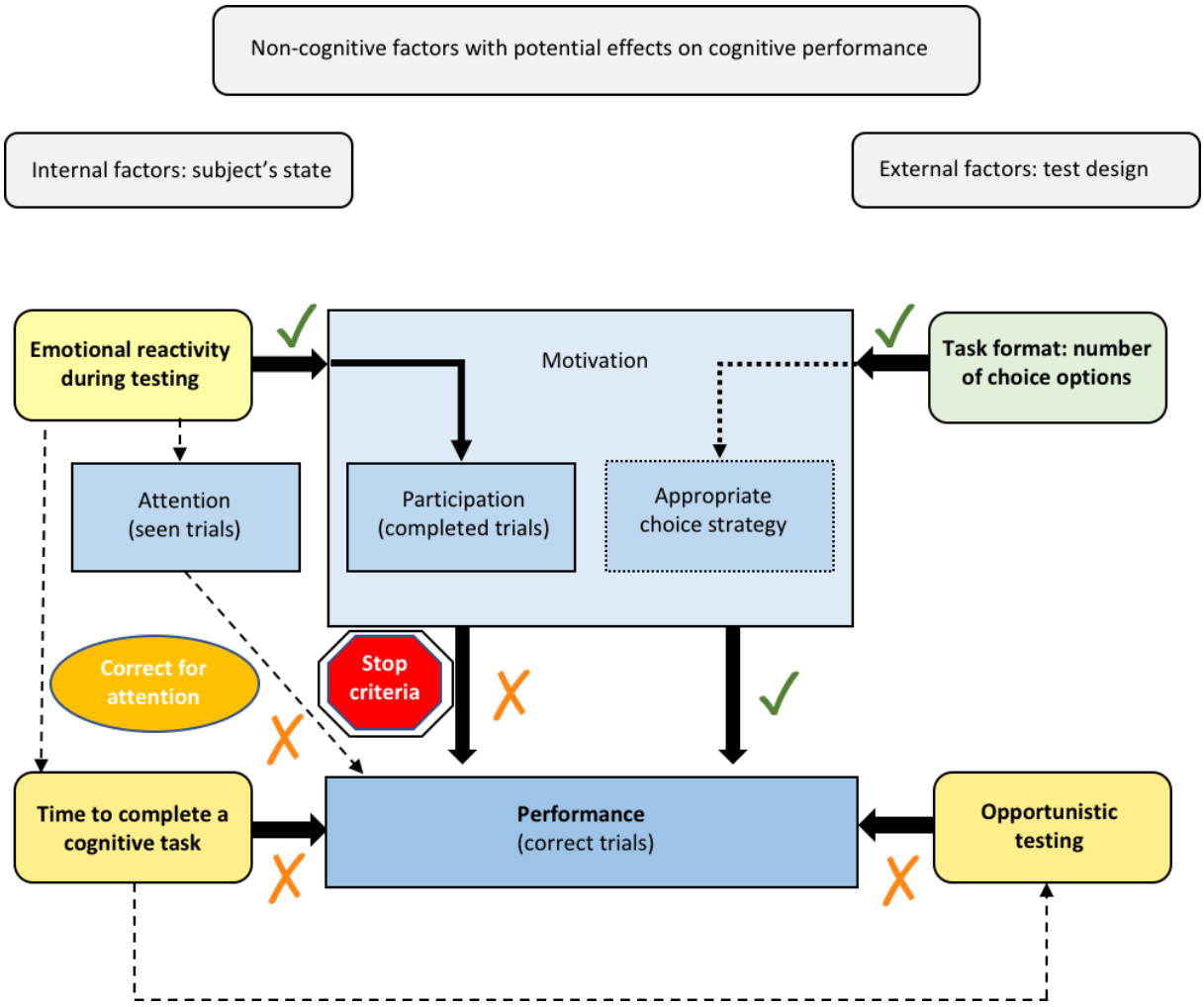


Figure 5.4 Summary of findings. Green ticks indicate that an effect of the respecting factor was found; orange crosses indicate no effect.

Chapter 6

Conclusion & Outlook

6.1 Contribution of my PhD thesis to comparative cognition in primates

The findings of this thesis add to the list of non-cognitive factors that have been studied so far regarding their potential to affect cognitive performance in physical cognition tasks (summarised in Table 6.1). My second and third study complement the external (test design/method-related) non-cognitive factors that have been addressed by other comparative cognition researchers working with non-human primates. While opportunistic testing and the selection bias that results from it did not affect performance in the two New World monkey species I tested, task format not only affected cognitive performance in one of my studies (and rendered the two-choice memory test version invalid) but it also did so in earlier studies conducted by other researchers. In fact, task format more generally can sometimes influence performance differently in different species. Two Old World primate species, for instance, olive baboons (*Papio anubis*) and long-tailed macaques (*Macaca fascicularis*), performed better in a quantity discrimination task (i.e., were more likely to choose the larger of two amounts of items), when the test stimuli were inedible rather than edible (Schmitt & Fischer, 2011). Interestingly, the monkeys performed equally well with edible test stimuli provided they were rewarded with the same number of food items they had chosen but of a different food type. The same experiment was later conducted with two New World monkey species (Gazes et al., 2018) and it was shown that brown capuchins (*Cebus sapajus apella*) performed best with edible rather than inedible items and outperformed squirrel monkeys (*Saimiri sciureus*) who performed relatively poorly overall. In a recent study, capuchins (*Sapajus spp.*) were tested with a visual object discrimination task in which either just visible or additional haptic information (via tactile exploration of the objects) was available. The capuchins performed better (i.e., took fewer trials to successfully distinguish between the two objects) when they were allowed to manipulate the objects by touch, thereby showing visuotactile integration (Carducci et al., 2018).

My findings (Study 1, *Chapter 2*) that individual differences in emotional reactivity affect the subjects' motivation to participate but not their cognitive performance add to the list of internal (subject-related) non-cognitive factors that potentially affect cognitive performance that have so far been quantified. Long-tailed macaques with trait anxiety (as assessed by a strong and stable behavioural reaction to a sudden loud noise) performed worse in a reversal learning task than their conspecifics without this trait (Toxopeus et al., 2005). In both

chimpanzees (*Pan troglodytes*) and orang-utans (*Pongo pygmaeus*), bolder subjects performed better in the physical subtests of a large cognitive test battery than shyer subjects (Herrmann et al., 2007). In the social subtests, however, temperament had no effect on the apes' cognitive performance. Certain personality traits, particularly openness and assertiveness, affected the participation and performance of capuchin monkeys (*Sapajus apella*) in a training task (Morton et al., 2013). More open and less assertive subjects participated and performed better than less open and highly assertive subjects. While temperament and personality traits are characterised through a high degree of stability over time, a subject's level of emotional reactivity may differ from situation to situation. The marmosets in my first study who showed a strong spontaneous emotional reaction to an experimenter and the test situation participated in fewer trials of an object permanence task than their less emotionally reactive conspecifics (Schubiger et al., 2015). Importantly, however, their cognitive performance was not affected.

Finally, rearing conditions appear to form a special class of factors. Per se they are external factors, but increasing evidence suggest they may causally change cognitive performance in primates (Damerius et al., 2017a, 2017b). For instance, rearing conditions affect primates' cognitive skill repertoires, most likely because having ample opportunity to learn socially from adult conspecifics (ideally by being raised by one's mother) allows primates to acquire a larger set of skills than being deprived of this opportunity (by growing up with peers as an orphan (reviewed in van Schaik & Burkart, 2011). In orangutans (*Pongo abelii* and *Pongo pygmaeus*), performance in a problem-solving task was best predicted by the degree of orientation towards humans (as measured by the Human Orientation Index (HOI)). Subjects who were more human-oriented as a consequence of their individual histories with humans were more explorative and more successful at using tools to retrieve honey from a wooden apparatus than less human-oriented subjects (Damerius et al., 2017a). Furthermore, being cared for by humans and being housed with conspecifics allowed orang-utans to be more curious and explorative than their single-housed conspecifics and perform better in several other physical cognition tests (e.g., inhibitory control and causal reasoning; Damerius et al., 2017b).

6.2 How can tests of primate cognition be improved?

I evaluated empirically, for the first time, the validity of a cognitive test battery for non-human primates. For this purpose, I used several criteria that we had proposed in an earlier publication (Burkart, Schubiger, and van Schaik, 2017a, see *Appendix*) to establish whether it truly measured individual differences in cognitive abilities (i.e., reasoning ability and behavioural flexibility) or whether a number of not primarily cognitive confounding factors affected task performance. My findings (summarised in Table 6.1) contribute to the recent efforts of improving existing cognitive tests that assess the cognitive abilities of non-human primates.

First, I experimentally controlled for emotional and motivational confounds in the first two studies (see *Chapters 2 & 3*). I found that marmosets express their emotional arousal levels (visibly and audibly) in their appearance and behaviour and tend to avoid the testing area once their arousal levels have peaked. Other species may express high arousal differently, but it is essential that the experimenter identifies behavioural indicators of arousal before and keeps monitoring them during cognitive testing.

Subjects experiencing above-average emotional arousal levels tend to be reluctant to fully participate in a cognitive task and get easily distracted when they do participate. In marmosets this was particularly the case for many male individuals who often tended to be vigilant towards their surroundings rather than the cognitive task and did often not complete a full test session on a given day. Although larger-bodied primate species with less predation risk may not need to be as emotionally reactive as the small-bodied marmosets, it is essential to use clear stop criteria for all non-human primate species. This ensures that testing is paused when a subject is clearly no longer motivated to participate in further trials which is essential for its wellbeing and to obtain meaningful performance scores.

Second, I empirically explored the role of the monkeys' motivation to perform well in cognitive tasks and improved the validity of a classical memory task by increasing the number of choice options which in turn motivated the monkeys to perform well (i.e., make as many correct choices as possible). Similar changes to the task format may also improve the validity of other cognitive tasks and species and reduce the probability that confounding factors such as motivational issues impact cognitive performance.

Taken together, with the exception of one of the memory tasks, I could rule out that the subtests of the cognitive test battery measured non-cognitive factors rather than true cognitive performance. Therefore, I empirically controlled for individual differences in the subjects' emotional reactivity (as quantified as the spontaneous emotional response to human experimenters in Study 1, *Chapter 2*) their motivation to participate (Study 1, *Chapter 2*) and perform well (Study 2, *Chapter 3*) in cognitive tasks.

In addition, while I prioritised testing the same individuals with all subtests and my sample size was therefore relatively small, the subtests consisted of diverse tasks tapping into several cognitive abilities: inhibitory control, understanding of quantities, cognitive flexibility, and long-term memory (Study 3, *Chapter 4*).

Third, I could replicate the results from Study 2 (*Chapter 3*) in two independent samples representing my study species and an evolutionarily closely related species. Both New World primate species performed better in the nine-choice than the two-choice memory test and only the former accurately measured long-term memory. Future studies should replicate in other species if this improvement of the task's internal validity also helps other species to perform better.

Fourth, comparable to general intelligence studies with humans, I avoided tasks that are prone to secondary modularisation, such as tool-use related ones, that would have measured cognitive abilities that appear to be evolutionary adapted (primary) modules but in fact need to be learned even by species with a high general cognitive ability (secondary modules). Instead, I ensured that all subtests were as abstract as possible (Studies 1-3, *Chapters 2-4*). Likewise, other studies should use sufficiently abstract test set ups to avoid measuring adapted abilities such as optimal food search strategies.

Table 6.1 Overview of the findings and recommendations for future primate cognition studies

Criteria to avoid statistical & methodological artefacts	Findings of this thesis	Recommendations for future studies
1) Empirical control for confounds: • Emotional reactivity ✓ Subject's spontaneous emotional reaction towards experimenter & test situation • Motivation ✓ Subject's readiness to participate, concentrate, and perform well in a cognitive test • Task format ✓ Chance probability of being successful	Study 1 (Chapter 2) ⇨ Subjects behaviourally expressed elevated emotional arousal (e.g., via persisting piloerection, vigilance & mobbing calls) and tended to avoid the test area ⇨ High emotional reactivity affected the subjects' motivation to participate in a cognitive test (i.e., how many test trials they completed) ⇨ Cognitive performance was not affected by emotional reactivity, if the experimenter used pre-defined criteria on when to stop a test session (i.e., no choice in 4 trials)	⇨ Assess behavioural signs of elevated emotional reactivity before and during testing ⇨ Define clear stop criteria before testing & apply them consistently (e.g., refusal to make a choice in *X* consecutive trials) ⇨ Apply these criteria to stop a test session if a subject shows signs of high reactivity or low motivation
	Study 2 (Chapter 3) ⇨ A subject's motivation to use an appropriate choice strategy was affected by task format (i.e., the number of options in a choice-task) in a memory test ⇨ A change to a cognitive task's presentation format (i.e., nine choice options made wrong choices costlier) improved the task's validity (i.e., the novel version measured the cognitive ability under question)	⇨ Be aware that motivational factors might bias performance, and design cognitive tests accordingly ⇨ Ensure the subjects' test scores capture individual differences in their cognitive abilities rather than in non-cognitive factors, such as motivational ones
2) Use of: • Diverse tasks ✓ • Analytical routines without an a priori categorisation of tasks into domains ✓ • Large samples ✗	Study 3 (Chapter 4) ⇨ The cognitive test battery consisted of six tasks from several broad domains of physical cognition: inhibition (2 tasks), quantity discrimination (1 task), cognitive flexibility (1 task), and memory (2 tasks)	⇨ Use a wider range of cognitive tests assessing various cognitive abilities ⇨ Use larger sample sizes if possible
3) Replication of results in independent samples, when large samples are not available ✓	Study 2 (Chapter 3) ⇨ Experimentally naïve marmosets & squirrel monkeys were tested with the same two memory test versions as the marmosets from the initial study sample (sample 1) ⇨ The poor performance results of the marmosets from sample 1 in the two-choice memory task version could be replicated with an independent marmoset sample (sample 2) and in a second species, squirrel monkeys ⇨ The better overall performance & decline with increasing memory demands in the nine-choice memory test was found in both species	⇨ Studies with different species are needed to further establish the external validity of the novel memory task
4) Avoidance of tasks prone to secondary modularisation ✓	Whole thesis; Studies 1-3 (Chapters 2-4) ⇨ In all cognitive tasks the test stimuli were presented in an abstract manner (e.g., a number of aligned containers) and no tasks required learned cognitive skills	⇨ Use abstract test set-ups, even when using as many as nine experimental stimuli, to avoid adapted search strategies

6.3 Is the test battery suitable for future studies on general intelligence?

The ultimate goal of this thesis was to determine if the cognitive tasks were suitable to be used in larger test batteries that could then be used in future studies to assess general intelligence (i.e., reasoning ability and behavioural flexibility) in large samples of non-human primates without risking that a psychometrically and statistically identified *g* would simply reflect a methodological and statistical artefact (Burkart, Schubiger & van Schaik, 2017a, b). While, apart from the two-choice memory task, the tasks may be suitable for general intelligence studies, using an opportunistic testing approach is not advisable.

Opportunistic testing, in which subjects with motivational issues are excluded from the study sample and replaced by more consistently motivated subjects, allows to minimise total testing time and to maximise sample sizes which can be advantageous for many cognitive studies. However, it still remains to be established whether opportunistic testing is as unproblematic for other cognitive tasks and other primate species. As my results suggest, it is not a favourable approach for general intelligence testing because it is important that every individual obtains a performance score in each task. Missing test scores result in small sample sizes because subjects with any missing test scores are excluded in principal component analyses used to identify the existence of *g*. Full testing is therefore advisable because, unlike in other cognitive studies, new subjects cannot simply be added to replace the dropouts as the study proceeds, even if they would be available. This is particularly the case if the test battery contains a larger number of cognitive tests, unless additional experimenters are used (which is justified if standardised methods are used, as my results have shown). A more favourable approach would be to conduct full testing but only use cognitive tests with a reasonable number of test sessions and trials. For instance, in the two tasks with the largest number of test sessions (Reversal Learning and Memory), some subjects eventually even dropped out when they were given several months to complete the task. This was particularly the case for the Reversal Learning task in which individual subjects were tested until they reached a learning criterion and indicates that those tasks that consist of a reasonably small number of test sessions reduce motivational issues and help to largely avoid the issue of dropouts.

6.4 Conclusion

In conclusion, this thesis contributes to the recent paradigm shift in comparative cognition research that acknowledges the importance of exploring individual variation within species. This shift in focus from comparing species to comparing individuals allows to better evaluate the measurement validity and reliability of cognitive tasks. While it is common practice in cognitive research with humans to evaluate cognitive test batteries regarding these quality criteria, such evaluations are currently missing in research with non-human primate species.

I took a first step into this direction by evaluating the role of non-cognitive confounding factors that might potentially affect cognitive performance and bias the results of cognitive studies. My findings demonstrate how researchers can evaluate and increase the validity of available cognitive tests, as shown for two New World primates and one cognitive test battery. If my findings generalise to other species, they will eventually allow fairer and more meaningful species comparisons that can help us to better understand our own cognitive evolution.

References

Agrillo, C. (2014). Numerical and arithmetic abilities in non-primate species. In: Kadosh, R. C., & Dowker, A. (Eds.), *The Oxford Handbook of Numerical Cognition*. Oxford, UK: Oxford University Press. doi:10.1093/oxfordhb/9780199642342.001.0001.

Amici, F., Aureli, F., & Call, J. (2008). Fission-fusion dynamics, behavioral flexibility, and inhibitory control in primates. *Current Biology*, *18*(18), 1415-1419. doi:10.1016/j.cub.2008.08.020.

Amici, F., Aureli, F., & Call, J. (2010). Monkeys and apes: are their cognitive skills really so different? *American Journal of Physical Anthropology*, *143*(2), 188-197. doi:10.1002/ajpa.21305.

Amici, F., Barney, B., Johnson, V.E., Call, J., & Aureli, F. (2012). A modular mind? A test using individual data from seven primate species. *PLoS One*, *7*(12), e51918.

Anderson, B. (1993). Evidence from the rat for a general factor that underlies cognitive performance and that relates to brain size: Intelligence? *Neuroscience Letters*, *153(*1), 98–102.

Anderson, J. R., Kuroshima, H., & Fujita, K. (2010). Delay of gratification in capuchin monkeys *(Cebus apella)* and squirrel monkeys *(Saimiri sciureus)*. *Journal of Comparative Psychology*, *124*, 205-210.

Arden, R., & Zietsch, B. (2017). An all-positive correlation matrix is not evidence of domain-general intelligence. *Behavioral and Brain Sciences*, *40*, e197. doi:10.1017/S0140525X16001990.

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The Rules of the Game Called Psychological Science. *Perspectives on Psychological Science, 7*(6), 543-554. doi: 10.1177/1745691612459060.

Banerjee, K., Chabris, C. F., Johnson, V. E., Lee, J. J., Tsao, F., & Hauser, M. D. (2009). General intelligence in another primate: individual differences across cognitive task performance in a New World monkey (*Saguinus oedipus*). *PLoS One, 4*(6), e5883. doi:10.1371/journal.pone.0005883.

Bassett L., Buchanan-Smith, H.M., McKinley, J., & Smith, T.E. (2003). Effects of training on stress-related behavior of the common marmoset (*Callithrix jacchus*) in relation to coping with routine husbandry procedures. *Journal of Applied Animal Welfare Science, 6*(3), 221–233.

Begun, D. R. (2015). *The real planet of the apes: a new story of human origins*. Princeton, NJ: Princeton University Press.

Bitterman, M. E. (1965). Phyletic differences in learning. *American Psychologist, 20*(6), 396-410. doi: 10.1037/h0022328.

Bitterman, M. E. (1988). Vertebrate-Invertebrate Comparisons. In: Jerison, H. J., & Jerison, I. (Eds.), *Intelligence and Evolutionary Biology*, vol 17. Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-70877-0_14.

Boyd, R., & Silk, J. B. (2012). *How humans evolved* (6th edition). New York: W.W. Norton.

Brown, J., Kaplan, G., Rogers, L. J., & Vallortigara, G. (2010). Perception of biological motion in common marmosets (*Callithrix jacchus*): by females only. *Animal Cognition, 13*(3), 555–564.

Burkart, J., & Heschl, A. (2006). Geometrical gaze following in common marmosets (*Callithrix jacchus*). *Journal of Comparative Psychology, 120*(2), 120-130. doi:10.1037/0735-7036.120.2.120.

Burkart, J. M., Fehr, E., Efferson, C., & van Schaik, C. P. (2007). Other regarding preferences in a nonhuman primate: common marmosets provision food altruistically. *PNAS, 104*(50), 19762–19766.

Burkart, J. M., Schubiger, M. N., & van Schaik, C. P. (2017a). The evolution of general intelligence. *Behavioral and Brain Sciences, 40*, e195. doi: 10.1017/S0140525X16000959.

Burkart, J. M., Schubiger, M. N., & van Schaik, C. P. (2017b). Future directions for studying the evolution of general intelligence. *Behavioral and Brain Sciences, 40*, e224. doi:10.1017/S0140525X17000024.

Burkart, J. M., Strasser, A., & Foglia, M. (2009). Trade-offs between social learning and individual innovativeness in common marmosets, *Callithrix jacchus. Animal Behaviour, 77*(5), 1291–1301.

Brown, G. R., Almond, R. E., & van Bergen, Y. (2004). Begging, stealing, and offering: food transfer in nonhuman primates. *Advances in the Study of Behavior, 34*, 265-295.

Call, J. (2006). Inferences by exclusion in the great apes: the effect of age and species. *Animal Cognition, 9*(4), 393-403. doi:10.1007/s10071-006-0037-4.

Campbell, D., T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin, 54* (4), 297-312.

Carducci, P., Schwing, R., Huber, L., & Truppa, V. (2018). Tactile information improves visual object discrimination in kea, *Nestor notabilis*, and capuchin monkeys, *Sapajus spp*. *Animal Behaviour, 135*, 199-207.

Carere, C., & Locurto, C. (2011). Interaction between animal personality and animal cognition. *Current Zoology, 57*(4), 491-498.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, UK: Cambridge University Press.

Clayton, N. S., & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature, 395*, 272-274.

Clune, J., Mouret, J. B. & Lipson, H. (2013) The evolutionary origins of modularity. *Proceedings of the Royal Society of London B: Biological Sciences, 280,* (1755), 20122863.

Cnotka, J., Güntürkün, O., Rehkämper, G., & Gray, D., & Hun, G. R. (2008). Extraordinary large brains in tool-using New Caledonian crows *(Corvus moneduloides). Neuroscience Letters, 433*(3), 241-245. doi: 10.1016/j.neulet.2008.01.026.

Cosmides L, & Tooby, J. (1992) Cognitive adaptions for social exchange. In: Barkow, J., Cosmides, L., & Tooby, J., (Eds.), The adapted mind: Evolutionary psychology and the generation of culture. New York: Oxford University Press.

Cosmides, L., & Tooby, J. (2002). Unraveling the enigma of human intelligence: Evolutionary psychology and the multimodular mind. *The evolution of intelligence*, 145-198.

Cosmides, L., & Tooby, J. (2013). Evolutionary psychology: New perspectives on cognition and motivation. *Annual Review of Psychology*, *64*, 201–229.

Cosmides, L., Tooby, J., Fiddick, L., & Bryant, G. A. (2005). Detecting cheaters. *Trends in Cognitive Sciences, 9*(11), 505-506.

Crofoot, M. C., Lambert, T. D., Kays, R., & Wikelski, M. C. (2010). Does watching a monkey change its behaviour? Quantifying observer effects in habituated wild primates using automated radiotelemetry. *Animal Behaviour, 80*(3), 475–480.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–300.

Damerius, L. A., Forss, S. I. F., Kosonen, Z. K., Willems, E. P., Burkart, J. M., et al. (2017a). Orientation toward humans predicts cognitive performance in orang-utans. *Scientific Reports, 7*, 40052. doi: 10.1038/srep40052.

Damerius, L. A., Graber, S., Willems, E. P., & van Schaik, C. P. (2017b). Curiosity boosts orang-utan problem-solving ability. *Animal Behaviour, 134*, 57-70. doi: 10.1016/j.anbehav.2017.10.005.

Deaner, R. O., Van Schaik, C. P., & Johnson, V. E. (2006). Do some taxa have better domain-general cognition than others? A meta-analysis of nonhuman primate studies. *Evolutionary Psycholo*gy, *4*(1), 149-196. doi: 10.1177/147470490600400114.

Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience, 11*(3), 201-211.

Dettling, A. C., Feldon, D., & Pryce, C. R. (2002). Repeated parental deprivation in the infant common marmoset *(Callithrix jacchus*, primates) and analysis of its effects on early development. *Biological Psychiatry, 52*, 1037–1046. doi:10.3758/BRM.41.2.396.

Ebbinghaus, H. (1885). *Über das Gedächtnis. Untersuchungen zur experimentellen Psychologie*. Leipzig: Duncker & Humber.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). New York: Columbia University, Teachers College. (Original work published 1885).

Emery, N. J., & Clayton, N. S. (2004). The mentality of crows: convergent evolution of intelligence in corvids and apes. *Science, 306*(5703), 1903-1907.

Emery, N. J. (2006). Cognitive ornithology: the evolution of avian intelligence. *Philosophical Transactions of the Royal Society B: Biological Science, 361*(1465), 23-43. doi: 10.1098/rstb.2005.1736.

Erdöhegyi, À., Topál, J., Virányi, Z., & Miklósi, Á. (2007). Dog-logic: inferential reasoning in a two-way choice task and its restricted use. *Animal Behaviour,* 74*,* 725–737. doi: 10.1016/j.anbehav.2007.03.004.

Ermer, E., Cosmides, L., & Tooby, J. (2007). Cheater detection mechanism. In R. F. Baumeister, R. F., & Vohs, K. D. (Eds.), *Encyclopedia of Social Psychology* (pp. 138-140). Thousand Oaks, CA: Sage Publications. doi: 10.4135/9781412956253.n82.

Fiddick, L., Spampinato, M., & Grafman, J. (2005) Social contracts and precautions activate different neurological systems: An fMRI investigation of deontic reasoning. *Neuroimage, 28*, 778–786.

Fischer, G. J., & Kitchener, S. L. (1965). Comparative learning in young gorillas and orangutans. *Journal of Genetic Psychology, 107*(2), 337-348.

Ford, S. M., Porter, L. M., & Davis, L. C. (2009). *The smallest anthropoids: the marmoset/callimico radiation* (developments in primatology: progress and prospects). New York: Springer.

Freitag, F. B., & Almeida Pessoa, D. M. (2012). Effect of luminosity on color discrimination of dichromatic marmosets (*Callithrix jacchus). Journal of the Optical Society of America A, 29*(2), A216-A222. doi: 10.1364/JOSAA.29.00A216.

French, G. M. (1959). Performance of squirrel monkeys on variants of delayed response. *Journal of Comparative and Physiological Psychology, 52*, 741-745.

Galsworthy, M. J., Paya-Cano, J. L., Liu, L., Monleón, S., Gregoryan, G., Fernandes, C., Schalkwyk, L. C., & Plomin, R. (2005). Assessing reliability, heritability and general cognitive ability in a battery of cognitive tasks for laboratory mice. *Behavioral Genetics, 35*(5), 675–692.

Galsworthy, M. J., Paya-Cano, J. L., Monleo, S. & Plomin, R. (2002). Evidence for general cognitive ability (g) in heterogeneous stock mice and an analysis of potential confounds. *Genes, Brain and Behavior, 1*, 88–95.

Gazes, R. P., Billas, A. R., & Schmitt, V. (2018). Impact of stimulus format and reward value on quantity discrimination in capuchin and squirrel monkeys. *Learning & Behavior, 46*(1), 89-100. doi: 10.3758/s13420-017-0295-9.

Geary, D. C., & Huffman, K. J. (2002) Brain and cognitive evolution: Forms of modularity and functions of mind. *Psychological Bulletin, 128*(5), 667. doi: 10.1037/0033-2909.128.5.667.

Gilad, Y., Wiebe, V., Przeworski, M., Lancet, D., & Pääbo, S. (2004). Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biology, 2*(1), e5. doi: 10.1371/journal.pbio.0020005.

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence, 24*,13–23.

Gray, D. T., Smith, A. C., Burke, S. N., Gazzaley, A., & Barnes, C. A. (2017). Attentional updating and monitoring and affective shifting are impacted independently by aging in macaque monkeys. *Behavioural Brain Research, 322*, 329-338.

Griffin, A. S., Guillette, L. M., & Healy, S. D. (2015). Cognition and personality: an analysis of an emerging field. *Trends in Ecology and Evolution, 30*(4), 207-214.

Güntürkün, O., & Bugnyar, T. (2016). Cognition without Cortex. *Trends in Cognitive Science, 20*(4), 291-303. doi: 10.1016/j.tics.2016.02.001.

Harlow, H. F. (1932). Comparative behavior of primates. III. Complicated delayed reaction tests on primates. *Journal of Comparative Psychology, 14*, 241–252.

Hayes, A. F., & Krippendorff, K. (2007) Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, *1*(1), 77–89.

Heldstab, S. A., Kosonen, Z. K., Koski, S. E., Burkart, J. M., van Schaik, C. P., & Isler, K. (2016). Manipulation complexity in primates coevolved with brain size and terrestriality. *Scientific Reports, 6*, 24528. doi:10.1038/srep24528.

Herndon, J. G., M. B. Moss, D. L. Rosene, & R. J. Killiany. (1997). Patterns of cognitive decline in aged rhesus monkeys. *Behavioural Brain Research, 87*, 25-34.

Herrmann, E., & Call, J. (2012). Are there geniuses among the apes? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 367*(1603), 2753-2761.

Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science, 317*, 1360–1366.

Herrmann, E., Hernández-Lloreda, M. V., Call, C., Hare, B., & Tomasello, M. (2010). The structure of individual differences in the cognitive abilities of children and chimpanzees. *Psychological Science*, *21*(1), 102-110.

Hopkins, W. D., Russell, J. L., & Schaeffer, J. (2014). Chimpanzee intelligence is heritable. *Current Biology, 24*(14), 1649-1652.

Huber, L. (2017). Where is the evidence for general intelligence in nonhuman animals? *Behavioral and Brain Sciences, 40*, e206. doi: 10.1017/S0140525X16001667.

Hunt, G. R. (1996). Manufacture and use of hook-tools by New Caledonian crows. *Nature, 379*, 249-251.

Hunt, G. R., & Gray, R. D. (2004). Direct observations of pandanus-tool manufacture and use by a New Caledonian crow *(Corvus moneduloides). Animal Cognition, 7*(2), 114-120.

Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika, 76*, 297–307.

Isden, J., Panayi, C., Dingle, C., & Madden, J. (2013). Performance in cognitive and problem-solving tasks in male spotted bowerbirds does not correlate with mating success. *Animal Behaviour, 86*(4), 829–838. doi: 10.1016/j.anbehav.2013.07.024.

Iwaniuk, A. N., Dean, K. M., & Nelson, J. E. (2005). Interspecific allometry of the brain and brain regions in parrots (psittaciformes): comparisons with other birds and primates. *Brain, Behavior and Evolution, 65*(1), 40-59.

Jarvis, E. D., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., et al. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews Neuroscience*, *6*, 151–159.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23, 524–532. doi: 10.1177/0956797611430953.

Joly, M., Micheletta, J., De Marco, A., Langermans, J. A., Sterck, E. H., & Waller, B. M. (2017). Comparing physical and social cognitive skills in macaque species with different degrees of social tolerance. *Proceedings of the Royal Society B: 284*, 1862, 20162738.

Keagy, J., Savard, J. F., & Borgia, G. (2011). Complex relationship between multiple measures of cognitive ability and male mating success in satin bowerbirds, *Ptilonorhynchus violaceus. Animal Behaviour, 81*(5), 1063–1070.

Kendrick, D. F., Rilling, M. E., & Denny, M. R. (1986). *Theories of animal memory*. Hillsdale, N.J: L. Erlbaum Associates.

King, G. E. (2016). *Primate Behavior and Human Origins*. London: Routledge.

Kittler, K., Kappeler, P. M., & Fichtel, C. (2018). Instrumental Problem-Solving Abilities in Three Lemur Species. (*Microcebus murinus, Varecia variegata*, and *Lemur catta*). *Journal of Comparative Psychology, 132*(3), 306-314. doi: 10.1037/com0000113.

Kovacs, K., & Conway, A. R. A. (2017). The evolution of fluid intelligence meets formative g. *Behavioral and Brain Sciences, 40,* e208. doi:10.1017/S0140525X16001680.

Koenig, A. (1998). Visual scanning by common marmosets *(Callithrix jacchus)*: functional aspects and the special role of adult males. *Primates, 39*(1), 85–90.

Kolata, S., Light, K., Grossman, H. C., Hale, G., & Matzel, L. D. (2007). Selective attention is a primary determinant of the relationship between working memory and general learning ability in outbred mice. *Learning & Memory, 14*(1–2), 22–28.

Kolata, S., Light, K., Townsend, D. A., Hale, G., Grossman, H. C., & Matzel, L. D. (2005). Variations in working memory capacity predict individual differences in general learning abilities among genetically diverse mice. *Neurobiology of Learning and Memory, 84*, 241–246.

Krippendorf, K. (2011). Computing Krippendorff's alpha-reliability. Retrieved from http://repository.upenn.edu/asc_papers/43.

Lacreuse, A., Russell, J. L., Hopkins, W. D., & Herndon, J. G. (2014). Cognitive and motor aging in female chimpanzees. *Neurobiolology of aging, 35*(3), 623-632.

Light, K. R., Kolata, S., Hale, G., Grossman, H., & Matzel, L. D. (2008) Up-regulation of exploratory tendencies does not enhance general learning abilities in juvenile or young-adult outbred mice. *Neurobiology of Learning and Memory, 90*, 317–329.

Light, K. R., Kolata, S., Wass, C., Denman-Brice, A., Zagalsky, R., & Matzel, L. D. (2010) Working memory training promotes general cognitive abilities in genetically heterogeneous mice. *Current Biology, 20*, 777–782.

Lind, J., Enquist, M., & Ghirlanda, S. (2015). Animal memory: A review of delayed matching-to-sample data. *Behavioural processes, 117*, 52-58.

Locurto, C. (2017). G and g: Two markers of a general cognitive ability, or none? *Behavioral and Brain Sciences*, 40, e211. doi:10.1017/S0140525X16001709, e211

MacLean, E. L., Hare, B., Nunn, C. L., Adessi, E., Amici, F., et al. (2014). The evolution of self-control. *PNAS, 111*(20), E2140-E2148. doi:10.1073/pnas.1324533111.

MacPhail, E. M. (1982). *Brain and Intelligence in Vertebrates*. Oxford: Clarendon Press.

MacPhail, E. M. (1987). The comparative psychology of intelligence. *Behavioral and Brain Sciences, 10*, 645–656.

Makel, M., Plucker, J., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science, 7*, 537–542.

Martins Bezerra, B., & Souto, A. (2008). Structure and usage of the vocal repertoire of *Callithrix jacchus. International Journal of Primatology, 29*, 671–701.

Mather, J. A., & Kuba, M. J. (2013) The cephalopod specialties: Complex nervous system, learning, and cognition. *Canadian Journal of Zoology, 91*, 31–49.

Mattison, J. A., & Vaughan, K. L. (2017). An overview of nonhuman primates in aging research. *Experimental Gerontolology, 94*, 41-45.

Matzel, L. D., Han, Y. R., Grossmann, H., Karnik, M. S., Patel, D., Scott, N., Specht, S. M., & Gandhi, C. C. (2003) Individual differences in the expression of a “general” learning ability in mice. *The Journal of Neuroscience, 23*(16), 6423–6433.

Matzel, L. D., Townsend, D. A., Grossman, H., Han, Y. R., Hale, G., Zappulla, M., Light, K., & Kolata, S. (2006). Exploration in outbred mice covaries with general learning abilities irrespective of stress reactivity, emotionality, and physical attributes. *Neurobiology of Learning and Memory, 86*, 228–240.

McCall, R. B., & Carriger, M. S. (1993). A meta-analysis of infant habituation and recognition memory performance as predictors of later IQ. *Child Development, 64*(1), 57-79.

Mendes, N., & Huber, L. (2004). Object permanence in common marmosets (*Callithrix jacchus*). *Journal of Comparative Psychology, 118*, 103–112.

Menzel, E., Juno, C., & Garrud, P. (1985). Social foraging in marmoset monkeys and the question of intelligence [and discussion]. *Philosophical Transactions of the Royal Society B*: *Biological Sciences, 308*(1135), 145-158.

Miklósi, A., Kubinyi, E., Topàl, J., Gácsi, M., Virányi, S., & Csányi, V. A. (2003). A simple reason for a big difference: wolves do not look back at humans, but dogs do. *Current Biology, 13*(9), 763-766. doi: 10.1016/S0960-9822(03)00263-X.

Mikolasch, S., Kotrschral, K., & Schloegl, C. (2011). Is caching the key to exclusion in corvids? The case of carrion crows *(Corvus corone corone). Animal Cognition, 15*, 73–82. doi: 10.1007/s10071-011-0434-1.

Miles, R. C. (1957a). Delayed-response learning in the marmoset and the macaque. *Journal of Comparative and Physiological Psychology, 50*, 352-355.

Miles, R. C. (1957b). Learning-set formation in the squirrel monkey. *Journal of Comparative and Physiological Psychology, 50*(4), 356-357.

Miles, R. C., & Meyer, D. R. (1956). Learning sets in marmosets. *Journal of Comparative and Physiological Psychology, 49*, 219-222.

Morton, F. B., Lee, P. C., & Buchanan-Smith, H. M. (2013). Taking personality selection bias seriously in animal cognition research: a case study in capuchin monkeys *(Sapajus apella)*. *Animal Cognition, 16*, 677–684. doi:10.007/s10071-013-0603-5.

Mulcahy, N. J., Call, J., & Dunbar, R. I. M. (2005). Gorillas (*Gorilla gorilla*) and orangutans *(Pongo pygmaeus)* encode relevant problem features in a tool-using task. *Journal of Comparative Psychology, 119*(1), 23-32.

Mulcahy, N. J., & Hedge, V. (2012). Are great apes tested with an abject object-choice task? *Animal Behaviour, 83*(2), 313-321. doi:10.1016/j.anbehav.2011.11.019.

Mulcahy, N. J., & Schubiger, M. N. (2014). Can orangutans (*Pongo abelii*) infer tool functionality? *Animal Cognition, 17*(3), 657-669. doi:10.1007/s10071-013-0697-9.

Newell, B. R., Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences, 37*, 1– 61. doi:10.1017/S0140525X12003214.

Pashler, H., & Wagenmaekers, E. J. (2012). A crisis of confidence? *Perspectives on Psychological Science, 7*(6), 528-530. doi: 10.1177/1745691612465253.

Pessoa, D. M. A, Tomaz, C., & Pessoa, V. F. (2005). Color vision in marmosets and tamarins: behavioural evidence, *American Journal of Primatology, 67*, 487-495.

Peres, C. A. (1992). Prey-capture benefits in a mixed-species group of Amazonian tamarins, *Saguinus fuscicollis* and *S. mystax*. *Behavioral Ecology and Sociobiology, 31*, 339–347. doi: 10.1007/BF00177774.

Petit, O., Dufour, V., Herrenschmidt, M., De Marco, A., Sterck, H. M., & Call, J. (2015). Inferences about food location in three cercopithecine species: an insight into the socioecological cognition of primates. *Animal Cognition, 18*(821). doi: 10.1007/s10071-015-0848-2.

Petter, M., Musolino, E., Roberts, W. A., & Cole, M. (2009). Can dogs *(Canis familiaris)* detect human deception? *Behavioural Processes, 82*, 109–118. doi: 10.1016/j.beproc.2009.07.002.

Pravosudov, V. V., & Roth II, T. C. (2013). Cognitive ecology of food hoarding: The evolution of spatial memory and the hippocampus. *Annual Review of Ecology, Evolution, and Systematics, 44*, 173–193.

Prior, H., Schwarz, A., & Güntürkün, H. (2008). Mirror-induced behavior in the magpie (*Pica pica*): evidence for self-recognition. *PLoS Biology, 6*, e202.

Ramus, F. (2017) General intelligence is an emerging property, not an evolutionary puzzle. *Behavioral and Brain Sciences, 40*, e217. doi:10.1017/S0140525X1600176X.

Reader, S. M., Hager, Y., & Laland, K. N. (2011). The evolution of primate general and cultural intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*(1567), 1017-1027. doi:10.1098/rstb.2010.0342.

Rodriguez, J. S., & Paule, M. G. (2009). Working memory delayed response tasks in monkeys. In: Buccafusco, J. J. (Ed.), *Methods of Behavior Analysis in Neuroscience* (2nd ed.). Boca Raton (FL): CRC Press/Taylor & Francis.

Rosati, A. G., Rodriguez, K., & Hare, B. (2014). The ecology of spatial memory in four lemur species. *Animal Cognition, 17*(4), 947-961. doi: 10.1007/s10071-014-0727-2.

Rosenthal, R. (1963). On the social psychology of the psychological experiment: the experimenter's hypothesis as unintended determinant of experimental results. *American Scientist, 51*(2), 268–283.

Rowe, N. (1996). The pictorial guide to the living primates. New York: Pogonias Press.

Rumbaugh, D. M. (1997): Competence, cortex and primate models: a comparative primate perspective. In: Krasnegor, N. A, Reid-Lyon, G., Golman-Racik, P. S. (Eds.), *Development of the prefrontal Cortex*. Baltimore: Brookes, pp. 117–140.

Rushmore, J., Leonhardt, S. D., & Drea, C. M. (2012). Sight or scent: Lemur sensory reliance in detecting food quality varies with feeding ecology. *PLoS One, 7*(8), e41558. doi: 10.1371/journal.pone.0041558.

Schellinck, H. M., Cyr, D. P., & Bron, R. E. (2010). How many ways can mouse behavioral experiments go wrong? Confounding variables in mouse models of neurodegenerative diseases and how to control them. *Advances in the Study of Behavior, 41*, 255–366.

Schmitt, V., & Fischer, J. (2011). Representational format determines numerical competence in monkeys. *Nature Communications, 2*(257). doi:10.1038/ncomms1262.

Schmonn, R. (2011). Action understanding and Theory of Mind in common marmosets *(Callithrix jacchus)*. Master thesis, University of Zurich, May 2011.

Schubiger, M. N., Kissling, A., Burkart, J. M. (2016). How task format affects cognitive performance: a memory test with two species of New World monkeys. *Animal Behaviour, 121*, 33-39.

Schubiger, M. N., Kissling, A., & Burkart, J. M. (2018). Does opportunistic testing bias cognitive performance in primates? Learning from drop-outs. *Manuscript submitted for publication*.

Schubiger, M. N., Wüstholz, F. L., Wunder, A., & Burkart, J. M. (2015). High emotional reactivity toward an experimenter affects participation, but not performance, in cognitive tests with common marmosets (*Callithrix jacchus*). *Animal Cognition, 18*(3), 701-712. doi:10.1007/s10071-015-0837-5.

Shaw, R. C., Boogert, N. J., Clayton, N. S., & Burns, K. C. (2015). Wild psychometrics: Evidence for "general" cognitive performance in wild New Zealand robins, *Petroica longipes*. *Animal Behaviour, 109*,101–11. doi: 10.1016/j.anbehav.2015.08.001

Sherry, D. F. (2006) Neuroecology. *Annual Review of Psychology, 57*, 167–197.

Shuker, D. M., Barrett, L., Dickins, T. E., Scott-Philipps T. C., & Barton, R. A. (2017). General intelligence does not help us understand cognitive evolution. *Behavioral and Brain Sciences, 40*, e218.

Siegwarth, N., Larkin, K. T., & Kemmner, C. (2012). Experimental effects on cardiovascular reactivity and task performance during mental stress testing. *Psychological Record, 62*(1), 69–81.

Sorge, R. E., Martin, L. J., Isbester, K. A., Sotocinal, S. G., et al. (2014). Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nature Methods, 11*, 629–632. doi:10.1038/nmeth.2935.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology, 15*(2), 201-292.

Spearman, C. (1927). *The abilities of man: their nature and measurement*. Oxford, England: Macmillan., xxii, 415pp.

Stevens, J. R. (2017). Replicability and Reproducibility in Comparative Psychology. *Frontiers in Psychology, 8*, 862. doi: 10.3389/fpsyg.2017.00862.

Stevens, J. R., Hallinan, E. V., & Hauser M. D. (2005). The ecology and evolution of patience in two New World monkeys. *Biology Letters, 1*(2), 223-226.

Stone, V. E., Cosmides, L., Tooby, J., Kroll, N., & Knight, R. T. (2002). Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. *PNAS, 99*(17), 11531-11536.

Strasser, A., & Burkart, J. M. (2012). Can we measure brain efficiency? An empirical test with common marmosets (*Callithrix jacchus*). *Brain, Behavior, and Evolution, 80*(1), 26-40. doi: 10.1159/000338014.

Sultan, F. (2005). Why some bird brains are larger than others. *Current Biology 15*(17), PR649-PR650. doi: 10.1016/j.cub.2005.08043.

Tomasello, M., & Call, J. (1997). *Primate Cognition*. New York: Oxford University Press.

Toxopeus, I. B., Sterck, E. H. M., van Hooff, J. A. R. A. M., Spruijt, B. M., & Heeren, T. J. (2005). Effects of trait anxiety on performance of sociallyhoused monkeys in a learning test. *Behaviour, 142*(9–10), 1269–1287. doi:10.1163/156853905774539373.

Treichler, F. (1964). Delayed-response performance by the squirrel monkey. *Psychonomic Science, 1*(1), 129-130. doi:10.3758/BF03342825.

Tricarico, E., Amodio, P., Ponte, G., & Fiorito, G. (2014). Cognition and recognition in the cephalopod mollusc Octopus vulgaris: coordinating interaction with environment and conspecifics. In: Witzany, G. (Ed.), *Biocommunication of Animals*. Berlin: Springer. doi: 10.1007/978-94-007-7414-8_19.

Truppa, V. (2018). Tactile information improves visual object discrimination in kea, *Nestor notabilis,* and capuchin monkeys, *Sapajus spp*. *Animal Behaviour, 135*, 199-207.

Tsujimoto, S., & Sawaguchi, T. (2002). Working memory of action: a comparative study of ability to selecting response based on previous action in New World monkeys *(Saimiri sciureus* and *Callithrix jacchus*). *Behavioural Processes, 58*(3), 149-155.

van Schaik, C. P., Isler, K., & Burkart, J. M. (2012). Explaining brain size variation: from social to cultural brain. *Trends in Cognitive Sciences, 16*(5), 277-284.

van Schaik, C. P., & Burkart, J. M. (2011). Social learning and evolution: The cultural intelligence hypothesis. *Philosophical Transactions of the Royal Society B, 366*, 1008-1016.

Voelkl, B., & Würbel, H. (2016) Reproducibility Crisis: Are We Ignoring Reaction Norms? *Trends in Pharmacological Sciences, 37*(7), doi: 10.1016/j.tips.2016.05.003.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12*, 129–140. doi: 10.1080/17470216008416717.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale-IV, technical and interpretative manual*. San Antonio, Texas: Pearson.

Wobber, V., Herrmann, E., Hare, B., Wrangham, R., & Tomasello, M. (2014). Differences in the early cognitive development of children and great apes. *Developmental Psychobiology, 56*(3), 547-573.

Woltereck, R. (1909). Weitere experimentelle Untersuchungen über Artveränderung, speziell über das Wesen quantitativer Artunterschiede bei Daphnien. *Verhandlungen der deutschen zoologischen Gesellschaft, 19*, 110–173.

Woodley of Menie, M. A., Fernandes, H. B., & Hopkins, W. D. (2015). The more g-loaded, the more heritable, evolvable, and phenotypically variable: Homology with humans in chimpanzee cognitive abilities. *Intelligence, 50*, 159-163.

Woodruff, G., & Premack, D. (1979). Intentional communication in the chimpanzee: The development of deception. *Cognition, 7*, 333-362.

Yamamoto, M. E., Domeniconi, C., & Box, H. (2004). Sex differences in common marmosets *(Callithrix jacchus)* in response to an unfamiliar food task. *Primates, 45*, 249–254.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*, 459–482. doi:10.1002/cne.920180503.

Appendix

Reprint of relevant co-authored publication

BEHAVIORAL AND BRAIN SCIENCES (2017), Page 1 of 67
doi:10.1017/S0140525X16000959, e195

The evolution of general intelligence

Judith M. Burkart
Department of Anthropology, University of Zurich, CH-8125 Zurich, Switzerland
Judith.Burkart@aim.uzh.ch
http://www.aim.uzh.ch/de/Members/seniorlecturers/judithburkart.html

Michèle N. Schubiger
Department of Anthropology, University of Zurich, CH-8125 Zurich, Switzerland
michele.schubiger@aim.uzh.ch
http://www.aim.uzh.ch/de/Members/phdstudents/micheleschubiger.html

Carel P. van Schaik
Department of Anthropology, University of Zurich, CH-8125 Zurich, Switzerland
vschaik@aim.uzh.ch
http://www.aim.uzh.ch/de/Members/profofinstitute/vanschaik.html

Abstract: The presence of general intelligence poses a major evolutionary puzzle, which has led to increased interest in its presence in nonhuman animals. The aim of this review is to critically evaluate this question and to explore the implications for current theories about the evolution of cognition. We first review domain-general and domain-specific accounts of human cognition in order to situate attempts to identify general intelligence in nonhuman animals. Recent studies are consistent with the presence of general intelligence in mammals (rodents and primates). However, the interpretation of a psychometric *g* factor as general intelligence needs to be validated, in particular in primates, and we propose a range of such tests. We then evaluate the implications of general intelligence in nonhuman animals for current theories about its evolution and find support for the cultural intelligence approach, which stresses the critical importance of social inputs during the ontogenetic construction of survival-relevant skills. The presence of general intelligence in nonhumans implies that modular abilities can arise in two ways, primarily through automatic development with fixed content and secondarily through learning and automatization with more variable content. The currently best-supported model, for humans and nonhuman vertebrates alike, thus construes the mind as a mix of skills based on primary and secondary modules. The relative importance of these two components is expected to vary widely among species, and we formulate tests to quantify their strength.

Keywords: brain size evolution; comparative approach; cultural intelligence; evolution of intelligence; general intelligence; modularity; nonhuman primates; positive manifold; psychometric intelligence; rodents; social learning; species comparisons

1. Domain-general and domain-specific accounts of human cognition

"Animal behavior is driven by instincts, whereas human beings behave rationally." Views like these are still commonly expressed and deeply anchored in the Western worldview (e.g., Pinker 2010). A modern version of this dichotomy construes animals as having domain-specific, modular cognitive adaptations, whereas humans have domain-general intelligence. However, we now know that in human cognition, domain-specific components are ubiquitous too (Cosmides & Tooby 2013), perhaps even in complex cognitive tasks such as logical inference (Cosmides et al. 2010) or solving Bayesian probability problems (Lesage et al. 2013). At the same time, much evidence has accumulated that nonhuman minds are not exclusively made up of domain-specific specializations, but that domain-general cognitive processes may also be widespread. These empirical findings have implications for contemporary theories of the evolution of general intelligence, highlighted in section 3, provided it is established that general intelligence in animals is both real and refers to the same construct as in humans.

JUDITH M. BURKART is a senior researcher at the Department of Anthropology of the University of Zurich. She leads the Evolutionary Cognition Group and is interested in the cognitive evolution of primates. A main focus of the group concerns the role of systematic allomaternal care, present in callitrichid monkeys and humans, in the evolution of social, motivational, and cognitive processes.

MICHÈLE N. SCHUBIGER is a Ph.D. student of the Evolutionary Cognition Group, University of Zurich, and investigates general intelligence in marmoset monkeys. Currently, she is also a fellow in psychology, primate cognition, at Abertay University, Dundee, studying cognitive abilities in gibbons.

CAREL P. VAN SCHAIK is Professor and Director of the Department of Anthropology, University of Zurich. His main interests are socioecology and social evolution in primates, especially the primate foundations of human culture and intelligence. He wrote *Among Orangutans: Red Apes and the Rise of Human Culture* (Belknap Press of Harvard University Press, 2004) and the recent textbook, *Primate Origins of Human Nature* (Wiley-Blackwell, 2016), and has co-edited various volumes on primate behavior and conservation.

 0140-525X/17

The evolution of general intelligence poses a major puzzle. Because modular systems may readily evolve (Pavlicev & Wagner 2012; Schlosser & Wagner 2004; Shettleworth 2012b; but see Anderson & Finlay 2014; Lefebvre 2014), the evolution of the mind as a set of domain-specific adaptations or modules can easily be imagined. Indeed, a small set of dedicated modules, without any domain-general cognitive abilities, to which additional modules can be added as needed, may be the ancestral state of vertebrate cognition. This perspective is so convincing that it has led to accounts of massive modularity, not only for animal cognition, but also for human cognition as well (reviews: Barrett 2015; Frankenhuis & Ploeger 2007; Hufendiek & Wild 2015).

Evolutionary pathways leading to the emergence of domain-general cognitive processes, on the other hand, may appear less straightforward, because such open-content processes translate far less reliably into fitness-enhancing behavior, and because they may also require disproportionate amounts of energetically costly brain tissue compared to domain-specific specializations (van Schaik et al. 2012). Consequently, compared to the evolution of additional cheap and reliable, domain-specific, specialized cognitive solutions to specific problems, the evolution of general cognitive processes might pose greater obstacles to natural selection. Nonetheless, humans possess general intelligence, and if general intelligence can also be found in nonhuman animals, we can attempt to identify the evolutionary processes that can lead to its emergence, including the specific case of humans.

The aims of this review are (1) to critically evaluate the evidence for general intelligence in nonhuman animals, and (2) to explore the implications of its presence in nonhumans for current theories of cognitive evolution. To achieve these aims, we will review the theoretical background and evidence from a variety of research traditions, such as animal behavior and psychology, psychometrics and developmental psychology, and evolutionary psychology. Whereas all of these fields share an interest in understanding how the mind works, they are not well integrated, and attempts at integration have not yet produced consensus (e.g., Eraña 2012; Evans 2011; 2013; Toates 2005). In this target article, we will therefore selectively focus on those aspects that are necessary to integrate the findings from animal studies on general intelligence with what is known about humans. As non-experts in several of these fields, we are aware that we may not fully represent all of the relevant aspects of the respective theories, let alone solve current controversies in individual fields. Nevertheless, we hope that this article serves as a first step in achieving the much-needed integration across these disciplines at a more fine-grained level, which will eventually enable the development of a more unified theory of cognitive evolution.

This article is structured as follows. We first briefly review conceptualizations of both domain-generality and domain specificity of human cognition, and use this as background to situate current evidence for general intelligence in nonhuman animals, which is increasingly reported in various species based on factor-analytical approaches. We examine alternative explanations for these findings and develop a set of empirical criteria to investigate to what extent a statistically derived psychometric factor does indeed correspond to general intelligence as broadly defined. Such criteria are increasingly met in rodent studies but are strikingly underexplored in primates or birds.

Next, we discuss different evolutionary theories that may explain why and how general intelligence can be widespread in nonhuman animals even though it is not immediately obvious how it can reliably produce fitness-enhancing behavior. We argue that the broad version of the cultural intelligence approach (Tomasello 1999; van Schaik & Burkart 2011; van Schaik et al. 2012) can best account for the current body of evidence. We end by proposing a model that construes the mind of both humans and nonhuman vertebrates as a mix of truly modular skills and seemingly modular skills that are ontogenetically constructed using general intelligence abilities. We refer to them as primary and secondary modules, respectively. Species differences are likely with regard to the importance of these components, and we formulate tests to quantify their strength.

1.1. The positive manifold and general intelligence

Intelligence in humans has been intensely studied for more than a century (e.g., reviewed in Deary et al. 2010; Nisbett et al. 2012). It is broadly defined as involving "the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is thus not merely book learning, a narrow academic skill, or test-taking smarts. Rather it reflects a broader and deeper capability for comprehending our surroundings – 'catching on,' 'making sense' of things, or 'figuring out what to do'" (Gottfredson 1997, p. 13). This definition has received broad acceptance (Nisbett et al. 2012). In animals, intelligence is thought to involve an individual's ability to acquire new knowledge from interactions with the physical or social environment, use this knowledge to organize effective behavior in both familiar and novel contexts, and engage with and solve novel problems (Byrne 1994; Rumbaugh & Washburn 2003; Yoerg 2001). Thus, general intelligence, as defined in either humans or nonhuman animals, stresses reasoning ability and behavioral flexibility.

The concept of human general intelligence is built on one of the most replicated findings in differential psychology. In humans, performance across tasks of different cognitive domains is positively correlated: the positive manifold. Factor-analytical procedures applied to large data sets of individual performance across tasks consistently reveal a single factor that loads positively overall and can explain a significant amount of variation, often termed *g* for (psychometric) general intelligence. Within this psychometric, factor-analytical approach, an individual's loading on this factor thus estimates its intelligence. Performance in specific cognitive tasks (e.g., Raven's Progressive Matrices) or test batteries (e.g., Wechsler Adult Intelligence Scale [WAIS]) is highly correlated with *g*, and is in fact often used as a proxy measure for it, for instance in studies aimed at localizing *g* in the brain (Burgess et al. 2011; Colom et al. 2006; Gläscher et al. 2010). In this article, we will speak of *general intelligence* when referring to the broad definition of Gottfredson (1997) that stresses reasoning ability and behavioral flexibility, and of *psychometric intelligence* when referring to the entity estimated by the psychometric variable *g*. For humans, it is generally assumed that *g* estimates general intelligence, based on the strong empirical correlations between the two, as reviewed below.

Psychometric intelligence, estimated through *g*, typically explains around 40% of variance in test performance, whereas the rest is explained by group factors and variance unique to specific tasks (Plomin 2001). It has been found that *g* in humans has a clear genetic foundation (Davies et al. 2011), and in the absence of adverse environments that compromise the complete deployment of an individual's capacity, heritability can explain remarkably high proportions of variance (Joshi et al. 2015; Nisbett et al. 2012). Furthermore, *g* has robust correlates in brain structure and function, such as brain size, gray matter substance, cortical thickness, or processing efficiency (Deary et al. 2010; Jung & Haier 2007). However, rather than being localized in specific brain parts, it seems to be a system-level property of the brain (Pietschnig et al. 2015). Finally, *g* is also a good predictor for various measures of life outcome, including school achievement, the probability of being in professional careers, occupational attainment, job performance, social mobility, and even health and survival. In particular, it is better at predicting such variables than specific cognitive abilities on their own (reviewed in Deary et al. 2010; Reeve 2004).

1.1.1. The structure of cognition. The structure of human cognition continues to be debated (e.g., Ortiz 2015). Nonetheless, the presence of *g* is now widely accepted due to the pervasive evidence from Carroll's (1993) seminal meta-analysis of over 460 carefully selected data sets on human cognitive ability. An influential account is Horn and Cattell's fluid-crystallized *gf-gc* model (see also Major et al. 2012). Fluid intelligence *gf* refers to the capacity to think logically and solve problems in novel situations independently of previously acquired knowledge, and to identify patterns and relationships, whereas crystallized intelligence *gc* refers to the ability to use skills, knowledge, and experience and crucially relies on accessing information from long-term memory. An explicit causal link from *gf* to *gc* is provided by investment theory (Cattell 1987), which is the developmental version of the *gf-gc* model and finds considerable empirical support (Thorsen et al. 2014).

An integrated version, the so-called CHC (Cattell-Horn-Carroll) theory, has been supported by several studies and is a widely accepted consensus model (McGrew 2009). The CHC model is hierarchical, placing a general factor *g* at the top, which affects both *gf* and *gc*. Most current models involve some hierarchical structure involving a general factor, *g*, and fluid intelligence, *gf* (but see, for instance, Bartholomew et al. 2009; Major et al. 2012; van der Maas et al. 2006). In fact, some have argued that *gf* and *g* represent the same entity (Kan et al. 2011), and the previously mentioned definition of intelligence in a broad sense in fact emphasizes elements of both constructs.

Some models of general intelligence that do not involve *g* are also still being considered. Van der Maas et al. (2006), for instance, have presented a dynamic model of general intelligence that assumes independent cognitive processes early in ontogeny. Over the course of development, the positive manifold emerges because of mutually beneficial interactions between these initially independent processes. To the extent that one agrees to equate general intelligence with the positive manifold, the mutualism model may be viewed as a model of general intelligence for human and nonhuman animals in which variation between species would reflect the extent to which mutually beneficial interactions between cognitive processes arise during development. Because, across species, bigger brains require more time to mature than smaller brains (Schuppli et al. 2012), and thus have more opportunities to develop such mutually beneficial interactions, such a scenario is compatible with an evolutionary perspective.

1.1.2. Executive functions and intelligence. Closely related to general intelligence are executive functions, or EFs (Barbey et al. 2012; Blair 2006). EFs refer to "general-purpose control mechanisms that modulate the operation of various cognitive subprocesses and thereby regulate the dynamics of human cognition" (Miyake et al. 2000, p. 50). In other words, they are "a family of top-down mental processes needed when you have to concentrate and pay attention, when going on automatic or relying on instinct or intuition would be ill-advised, insufficient, or impossible" (Diamond 2013, p. 136). Three core EFs can be distinguished, namely inhibitory control (behavioral inhibition, cognitive inhibition, and selective attention), working memory (Baddeley 2010), and cognitive flexibility.

Various measures of EFs have shown strong correlations with *g*/*gf*. Whereas the average correlation between working memory and *g* is 0.72, in some studies using latent variable analysis, it even reached identity (Colom et al. 2005; Nisbett et al. 2012), leading some authors to suggest that the two cannot be distinguished from each other (Royall & Palmer 2014). That *g* and EF are closely related is consistent with two further lines of evidence. First, working memory can be trained, and these training gains can translate into gains in general intelligence even though not all procedures are effective, and it is not always clear whether the training affects working memory per se or instead improves learning strategies (reviewed in Klingberg 2010; Morrison & Chein 2011; Nisbett et al. 2012; Shipstead et al. 2012). Second, growing up bilingually, which makes high demands on a variety of EFs on a routine basis, is associated with stronger EFs in non-linguistic contexts, and thus with *g* (Abutalebi & Clahsen 2015; Bialystok et al. 2012; Rabipour & Raz 2012). Nonetheless, because EFs do not provide the logical problem-solving functions and learning that are the hallmark of general intelligence (Embretson 1995), some aspects of general intelligence are independent of EFs.

In sum, evidence for domain-general intelligence in humans, estimated by the first factor derived in psychometric, factor-analytical approaches, is pervasive, and is backed up by neurobiological evidence and various correlates of life-outcome measures. The psychometrically derived *g* factor is thus consistent with the broad notion of general intelligence, which stresses reasoning ability and behavioral flexibility and invokes cognitive processes such as learning and remembering, planning, and executive functions. This conclusion raises the question of the evolutionary origin of general intelligence in humans, which we will address by reviewing recent developments in the nonhuman literature. To do so, we will review evidence for *g* in animals, and whether it is warranted to assume that *g* in animals is also consistent with a broader notion of general intelligence.

Intelligent behavior needs to be distinguished from behavior that may appear intelligent but lacks flexibility (Shettleworth 2012a). Intelligent behavior in animals is often referred to as behavior that shows some degree of flexibility and emanates from some kind of mental

representation rather than immediate perception only (Tomasello & Call 1997). For instance, when digging wasps are interrupted anywhere in the sequence of actions involved in measuring the size of a hole to place a larva together with a prey item into it, they must start again at the very beginning of the behavioral sequence (Wooldridge 1968). Thus, many behaviors that at first sight look like they are the product of reasoning or learning turn out to be inflexible adaptations or modules (Sherry 2006). A collection of such dedicated adaptations presumably represents the ancestral state (e.g., Shettleworth 2012a; 2012b), and thus the null model against which the hypothesis of general intelligence has to be tested. Before turning to nonhuman animals, we will therefore provide an overview of domain-specific, modular conceptions of the mind that have been put forward particularly, but not exclusively, by evolutionary psychologists.

1.2. Cognitive adaptations and domain specificity

A domain-general factor of intelligence can be contrasted with domain-specific cognitive mechanisms or adaptive specializations (Cosmides & Tooby 2002). The basic idea is that whenever a fitness-relevant cognitive problem arises repeatedly and predictably over long periods of time in a given species, natural selection favors a genetically based, developmentally canalized ("hardwired") solution to this problem. For instance, natural selection may provide a species with a particularly strong spatial memory to retrieve stored food, without endowing it with more-powerful cognitive capacities in other contexts (Sherry 2006). Importantly, domain-specific mechanisms cannot be used in domains other than the ones for which they evolved, whereas domain-general mechanisms can be used to solve problems across domains.

Thus, the mind of animals, including humans, can be conceived of as a collection of adaptive specializations, often construed as modules, each of which evolved to solve a specific adaptive problem (Duchaine et al. 2001). Notice that a mind uniquely made up of these kinds of specific adaptations is arguably incompatible with standard accounts of intelligence, because virtually no learning and flexibility are involved. Similarly, none of these specific cognitive adaptations require the presence of the domain-general processes underlying intelligence such as executive functions.

1.2.1. Modularity and general intelligence. A modular organization of mind is particularly appealing to evolutionary thinking because modular systems allow parts to be removed, added, or modified without affecting the function of the structure as a whole. Therefore, modular systems may be more evolvable or even the only evolvable systems (Clune et al. 2013; Pavlicev & Wagner 2012; Ploeger & Galis 2011; Schlosser & Wagner 2004; Shettleworth 2012b). Thus, whenever conditions are sufficiently stable or at least predictable across generations, natural selection should favor solving recurrent fitness problems via modules rather than via general cognitive processes, because the former solve these problems on average quickly, effortlessly, and efficiently (Cosmides et al. 2010) and can presumably evolve more readily. General intelligence, in contrast, is thus expected to evolve under conditions of social or environmental unpredictability. Solutions to these evolutionarily novel problems have to be acquired effortfully, via slow learning (e.g., Geary 2005; Geary & Huffman 2002).

The advantages of a modular solution to recurrent fitness problems, however, are not necessarily as straightforward. First, the fundamental assumption that a modular solution is indeed more evolvable can be questioned on both empirical and conceptual grounds (e.g., Anderson & Finlay 2014; Bolhuis et al. 2011; d'Souza & Karmiloff-Smith 2011; Lefebvre 2014). Empirical evidence for a direct mapping of specialized adaptive behavioral functions to specific modular neural units is actually rare, even for neural systems as simple as those of invertebrates. Novel adaptive functions seem mostly to be achieved via massive re-use of neural tissue rather than via the addition of encapsulated neuronal pools. Conceptually, the evolvability argument seems largely incompatible with what is known about short-term neuromodulation, brain plasticity over the life span, response to damage, and ontogenetic principles of brain development. The a priori evolvability argument, therefore, does not lead to an unambiguous conclusion as to the superiority of domain-specific over domain-general organization.

Second, the other advantage of modularity – fast, effortless, and ultimately efficient solving of evolutionarily recurrent fitness problems – may hold only for particular notions of modularity, such as Fodorian modules (Fodor 1983). These are thought to be domain-specific functional units that process distinctive input stimuli using distinctive mechanisms. In particular, a module is thought to exclusively process information from a specific domain and to produce a correspondingly specific output in the form of representations and/or a behavioral response. Fodor listed criteria that must – at least to "some interesting extent" (Fodor 1983, p. 37) – be fulfilled by a functional unit to qualify as modular. These criteria include domain specificity, mandatory processing, high speed, production of shallow outputs (i.e., not requiring extensive processing), limited accessibility, a characteristic ontogeny (reliable emergence without explicit learning), a fixed neural architecture, and informational encapsulation (meaning it is not affected by other cognitive processes, a criterion thought to be particularly important). Paradigmatic examples of Fodorian modules are optical illusions. Accordingly, the presence of modules involving the processing of sensory information is widely accepted, and that their speed and efficiency are beneficial is obvious. However, a modular organization has also been proposed for more higher-level cognitive processes including ones related to folk psychology (e.g., processing of faces and facial expressions, theory of mind, cheater detection), folk biology (e.g., animate-inanimate distinction, flora-fauna), or folk physics (e.g., movement trajectories, gravity biases, representation of space, solidity, and causality; summarized in Geary 2005). Indeed, massive modularity accounts hold that the mind is exclusively made up of modules (Barrett 2015; Carruthers 2005; Sperber 2001).

Massive modularity would appear to be irreconcilable with general intelligence (and therefore with the ability to solve evolutionarily novel problems), but much of the long-standing controversy about the massive modularity hypothesis of the human mind comes down to the use of different notions of modularity (see also Barrett & Kurzban 2006). Indeed, a variety of highly divergent notions have developed (Barrett 2015; Barrett & Kurzban 2012; Chiappe &

Gardner 2012; Coltheart 2011; Grossi 2014; Mahon & Cantlon 2011), and many of these are much broader than the Fodorian one (e.g., Sternberg 2011). Because they also encompass the possibility of overarching, central control processes (Carruthers 2011), they are entirely compatible with the coexistence of domain-general processes and general intelligence (Barrett 2015; Carruthers 2011). In fact, Carruthers (2011) argued that most modules are specialized learning systems. Such broad notions of modularity, however, arguably no longer support the original idea of automatically providing fast and frugal solutions to recurrent fitness problems.

Unlike many proponents of massive modularity in humans, comparative behavioral biologists and comparative psychologists typically refer to notions of modularity that hew closely to the classical Fodorian modules, that is, dedicated, inflexible cognitive adaptations that have evolved in response to specific recurrent fitness-relevant problems (e.g., Fernandes et al. 2014; Shettleworth 2012a; 2012b). Functional specialization here is mostly used in the biological, ultimate sense – that is, referring to the specific adaptive pressures that gave rise to the evolution of specific dedicated modules. This perspective is grounded in research traditions such as neuroecology (Sherry 2006) that have provided empirical evidence for the occurrence among animals of dedicated cognitive adaptations, such as spatio-temporal memory abilities in food-caching species, birds in particular (Brodin 2010; Pravosudov & Roth 2013). These cognitive adaptations typically do not generalize to problems for which they did not evolve.

A mind composed of such dedicated adaptations represents a plausible null model, and indeed a plausible ancestral state of vertebrate cognition. Dedicated adaptations and general intelligence can obviously coexist (e.g., Cosmides et al. 2010; Geary 2005) – for instance, when the output of modules serve as inputs for intelligent reasoning, which may be responsible for the fact that in humans general intelligence predicts reasoning ability even in evolutionarily familiar contexts (Kaufman et al. 2011). The key questions with respect to the evolution of general intelligence, therefore, are how central, domain-general processes could evolve on top of domain-specific adaptations, whether and to what extent they also exist in nonhuman animals, and what adaptive benefits drove their evolution.

1.2.2. Adaptive canalization beyond modularity. Strictly domain-general approaches that construe the mind as a general-purpose computer face several well-known problems (Cosmides & Tooby 1994; Cosmides et al. 2010; Frankenhuis & Ploeger 2007; Heyes 2003; Kolodny et al. 2015; see also Table 1). First, an agent has to efficiently identify relevant information and filter out irrelevant information in the process of problem solving, a challenge known as the frame problem. Second, once the relevant information has been identified, the agent has to decide what to do with it. To do so, she has to solve the problem of how to pick and combine correct, adaptive behavioral options or cognitive processes out of an exponentially growing number of possibilities (the problem of the combinatorial explosion) or to learn important associations and skills in a limited period of time despite dealing with relevant stimuli that occur at a low rate (the poverty of the stimulus problem). Third, correct responses have to be made quickly and efficiently (the urgency problem). And fourth, while doing so, the agent has to find general, rather than only locally successful, solutions (the functionality problem). It is thus beyond doubt that some canalization of cognitive processes is necessary.

Evolved Fodorian modules (referred to as "cognitive adaptations" by behavioral biologists and neuroecologists) are clearly one way of solving the problems highlighted previously, in particular when they define the entire sequence from the acquisition of information to the adaptive behavioral response. However, they are not necessarily the only possible way, and natural selection may also overcome these problems in a different way that would allow domain-general abilities to evolve. A straightforward solution to this problem would be that domain-general abilities coevolve together with adaptive canalizing mechanisms that guide how general abilities are applied. Canalizing mechanisms can have a phylogenetic origin, such as a genetically predetermined preference for a certain category of stimuli: for example, the preference for faces in human infants (Shah et al. 2015). Alternatively, they can have an ontogenetic origin, such as the propensity of chimpanzees from tool-using communities to automatically perceive a stick as a potential tool, compared to genetically indistinguishable chimpanzees from non-tool-using communities who do not recognize this affordance (e.g., Gruber et al. 2011).

Table 1 summarizes the phylogenetic and ontogenetic canalizing mechanisms that ensure that domain-general cognition produces adaptive behavior despite the problems highlighted previously. Unlike Fodorian modules, these mechanisms do not define the entire sequence from signal detection to behavioral output, but may be deployed at different stages during information processing. We will now examine the evidence for such domain-general canalization processes.

The first problem an individual faces is what to attend to in the continuous stream of stimuli coming in from different sensory modalities. This can be solved by innate dispositions or data acquisition mechanisms (also referred to as phylogenetic inflection: Heyes 2003). Importantly here, innateness is not equivalent to inflexibility because innate dispositions to pay attention to one stimulus over another can be conditional. For instance, an animal foraging for berries may have an attentional bias to perceive small red entities, but the same animal when exposed to a raptor will be biased to perceive only potential hideouts. Alternatively, animals can learn ontogenetically which targets are particularly worth attending to (ontogenetic inflection). Here, social guidance of attention may play a particularly important role. Ontogenetic inflection automatically arises whenever immatures follow the mother and later other conspecifics, and is even more powerful in species that follow gaze (Shepherd 2010). In many species, including humans, immatures are particularly attracted to everything conspecifics are interacting with, and immatures of some species, such as aye-ayes (Krakauer 2005), marmoset monkeys (Voelkl et al. 2006), or orangutans (Forss et al. 2015) are highly neophobic toward stimuli they have not witnessed their mother or other familiar conspecifics interact with. Natural selection can, therefore, favor the disposition to preferentially use social information to decide which stimuli to attend to, and thus leave the specific target of attention largely unspecified.

In a second step, the individual has to "decide" what to do with the stimuli that have captured its attention,

Table 1. *Overview of some specific problems that a domain-general cognitive apparatus has to overcome in order to produce ultimately adaptive behavior, as well as potential solutions – that is, adaptive canalization mechanisms. Note that these solutions may be very general themselves, such as a preference for social learning. See text for references.*

Problem	Domain-General Canalization Processes	Examples
The frame problem: *What to attend to?*	Input filters (phylogenetic inflection)	Facilitated detection of small red entities (when hungry) or dark openings (when chased)
	Socially guided attention (ontogenetic inflection)	Immatures following mothers, or following mothers' gaze
Problems of combinatorial explosion and poverty of stimulus *What to do with the information?*	Direct triggering, prepared learning	Flight reactions, learning to be fearful of snakes but not flowers
	Socially guided learning	Copying how to extract food from a matrix
	Integration with core knowledge[1]	Embedding the expectation that objects always fall down in a straight line (gravity bias) with knowledge of solidity
The urgency problem: *How to reach a quick, efficient response?*	Innate response tendencies	Evolved modules, evolved heuristics (primary modules)
	Acquired response tendencies (automatization, secondary modules)	Learned heuristics to solve algebraic equations (secondary modules)
The functionality problem: *How to find generally, not only locally, successful solutions?*	Innate goals	Innate template of a safe burrow, or of good food
	Socially acquired end-state preferences	Learning by following mother what a good sleeping place is; copying the goals of successful individuals, conformity biases

[1]That is, evolved cognitive domains that are fleshed out with experience; for example, Gelman (1990).

because input mechanisms filter incoming stimuli but do not produce behavior. Subsequent processes are therefore required to determine what to do with these stimuli without being stymied by the problems of poverty of stimulus and the combinatorial explosion. First, in the case of phylogenetic inflection, coevolution of input mechanisms and response tendencies is frequent (Lotem & Halpern 2012), as when a moving stimulus in the sky automatically triggers a flight reaction, but also when individuals are more likely to associate a snake (but not a flower) with fear (Cook & Mineka 1989), or a taste (but not an auditory stimulus) with subsequent nausea (known as biologically prepared learning or the Garcia effect: Garcia & Koelling 1966). Second, in the case of ontogenetic inflection, social learning can also affect how the individual processes a stimulus that has come to its attention. Third, the stimuli that have attracted an individual's attention may be integrated with innate bodies of knowledge, so-called core knowledge (Gelman 1990; Spelke & Kinzler 2007) or psychological primitives (Samuels 2004), and so give rise to more elaborate skills and conceptual systems (Carey 2009).

A third problem for the individual is that decisions often have to be made under time pressure (the urgency problem). Evolved modules, heuristics, or direct and reflexive triggering of responses are particularly good at providing fast responses because they bypass central processes. But quick and efficient responses can also be achieved in evolutionarily novel contexts, such as solving algebraic equations or playing chess, if a learned heuristic approach becomes an automated subroutine and can be applied effortlessly (Bilalić et al. 2011; Chang 2014). Such problem solving has similar surface properties to modular organization *sensu* Fodor. This fact has sometimes led to conceptual misunderstandings (see also section 1.2.3), and is relevant for approaches that try to identify domain-general processes in nonhuman animals (see also section 2.4.3).

A final potential problem is that developmentally acquired response tendencies may be successful in solving local problems, but nevertheless may not ultimately help an individual survive and reproduce (the functionality problem). Individuals, be they animals or humans, typically do not represent ultimate fitness goals in their everyday behavior. Rather, they pursue a set of innate psychological goals, which on average results in fitness-enhancing behaviors (Tinbergen 1963) but may become maladaptive in environments other than the one in which the goals evolved, as shown by our strong preferences for sweet, fatty, and salty foods. However, innate goals may be modified or supplemented by socially acquired end-state preferences. For immatures, who are most strongly affected by the canalization problems listed in Table 1, copying successful adult individuals is widespread and generally results in adaptive behavior because they are copying individuals who have survived until adulthood and managed to reproduce. Socially acquired end-state preferences and goals are particularly widespread in humans, who are highly susceptible to conformity and prestige biases (Dean et al. 2014; Richerson et al. 2016). Increasing evidence also suggests the existence of such biases in at least some nonhuman primates and birds (Aplin et al. 2015; Kendal et al. 2015; Luncz & Boesch 2014; van de Waal et al. 2013).

Despite being incomplete, Table 1 serves to highlight that adaptive canalization of cognition not involving Fodorian modules is possible, indeed potentially quite frequent. It also highlights the prominent reliance on social inputs to overcome the canalization problems inherent to domain-general mechanisms. Social learning is broadly defined in the animal literature – that is, learning influenced by observation of, or interaction with, another animal or its products (Heyes 1994; see also Box 1984). It is widespread in the animal kingdom, both in vertebrates and invertebrates, and ranges from processes as simple as social facilitation and enhancement learning to observational forms of social learning such as true imitation (e.g., Hoppitt & Laland 2013). Interestingly, it is increasingly assumed that many of the cognitive mechanisms involved in social learning are of a general nature rather than specialized, and are thus not specific to social learning (Behrens et al. 2008; Heyes 2012; 2016). Indeed, all forms of social learning also include a major element of individual learning. This is most evident in forms such as stimulus enhancement, where the attention of a naïve individual is drawn to stimuli other individuals are interacting with, which then releases individual exploration, play, and trial-and-error learning with this stimulus. Individual learning and practice, however, are also involved in the acquisition of skills through imitation learning, whereby it is typical that, after observation, a phase of individual practice is required (Galef 2015; Jaeggi et al. 2010; Schuppli et al. 2016). Thus, natural selection for social learning seems to automatically trigger selection on individual learning and general cognitive ability, suggesting that ontogenetic canalization through social learning may have contributed to enabling the evolution of domain-general cognition, an issue to which we return in section 3.3.

1.2.3. Primary and secondary modularization, and implications for general intelligence in nonhuman animals. Evolved Fodorian modules have specific surface properties: they work fast, effortlessly, and automatically, and they do not require significant amounts of executive control and working memory. Nevertheless, identifying modules in animals based on these properties is problematic because skills, capabilities, and solutions to problems that are acquired through effortful problem solving and learning based on general cognitive processes may become automatized over time, a process we refer to as secondary modularization. After such secondary modularization, or automatization, these skills have many of the surface properties in common with primary, evolved Fodorian modules. Note that this distinction in primary and secondary modularization is analogous to the distinction in primary and secondary cognitive abilities by Geary (1995), but whereas the latter has been developed specifically for humans, the former is thought to apply to a broad array of animal species.

Despite the similarities in surface properties, primary and secondary modules differ fundamentally with regard to their origin (see Table 2): Primary modules are evolved adaptations with canalized, buffered development, whereas secondary modules represent ontogenetically acquired skills that were automatized during ontogeny. In fact, secondary modularization is particularly common during the immature period (d'Souza & Karmiloff-Smith 2011). A consequence of the different etiology of primary and secondary modules is that the latter are more variable in their content and distribution across individuals or populations of the same species. Because little is known about the ontogeny of many of the specialized cognitive modules postulated for humans (Geary 2005), we should also acknowledge the possibility that some or all of these are secondary rather than primary (Anderson & Finlay 2014) or at least subject to experiential influences. For instance, even some prototypical modules such as those involved in face perception depend on experience (Dahl et al. 2014).

The implication for the question of general intelligence in nonhuman animals is that it is no longer possible to uniquely rely on surface properties such as speed, effort, efficiency, and reliability to infer the presence of evolved domain-specific modules, because secondary modules have similar properties. Instead, a better diagnostic tool

Table 2. *Primary and secondary modules differ with regard to their etiology and development, which has implications for their content and distribution within a species or population*

Type of Module	Etiology	Development	Content of Skills	Distribution	Examples
Primary modules	Evolutionary; reflect natural selection for domain-specific cognitive adaptation	Skill matures, motor practice (experience-expectant[1])	Preset, highly predictable	Uniformly present in a given species	Tendency of (young) felids to respond to small moving objects with behaviors from the hunting repertoire
Secondary modules	Ontogenetic; reflect behavioral flexibility and learning ability, acquisition often based on EFs	Skill is learned (experience-dependent[1]) and practiced to the point of automaticity	More variable, determined by nature of inputs	Variable among individuals, populations	Automatic perception of a stick as potential tool in some apes; learned algorithms to solve algebraic equations in humans

[1]Greenough et al. (1987).

for the presence of general cognitive abilities is the presence of variable skill profiles across individuals and genetically similar populations due to secondary modularization (see section 2.4.3).

We have shown that human cognition involves elements of domain-specific and domain-general processes, but that the same can potentially be true for animals as well. Hence, animal minds need not be bundles of specialized cognitive adaptations. Having thus leveled the playing field, we first, in section 2, review recent evidence for whether a positive manifold (*g*) is present in nonhuman animals at all, and if so, how such a *g* factor is best explained. In particular, we will focus on the question whether such psychometric intelligence shows any of the features usually referred to as general intelligence. Even if we can be confident that this is the case in humans, whether the same applies to animals must be an empirical question (Galsworthy et al. 2014), and we highlight different research strategies that may prove to be fruitful in the future. In section 3, we then use this pattern of results to examine the ultimate evolutionary question of why general intelligence evolved, and which selection pressures may have favored it.

2. General intelligence in nonhuman animals?

Unless general intelligence is inextricably linked to language, considerations of evolutionary continuity suggest that nonhuman animals, especially our closest extant relatives, the great apes, may well possess it too, at least to some extent. The presence of evidence for executive functions in animals (Chudasama 2011) supports this contention, as does the overall flexibility of brains in animals, both during development and as response to experience, including the training of cognitive skills (Johansen-Berg 2007; Kolb & Gibb 2015; Matsunaga et al. 2015; Sale et al. 2014). According to most neurobiologists, such developmental plasticity is incompatible with purely domain-specific descriptions of cognitive abilities (Anderson & Finlay 2014; Prinz 2006; Quartz 2003). Nonetheless, evolutionary plausibility does not amount to empirical evidence, to which we turn now.

The question of whether general intelligence is unique to humans has typically been addressed by asking whether we find a positive manifold or psychometric intelligence, by following two complementary approaches: First, within a given species, in analogy to human studies, psychometric test batteries have been applied to many individuals. Second, broad comparative analyses (both experimental and meta-analytical) have been conducted across species to investigate whether species differ from each other in general intelligence, rather than in specific cognitive adaptations. In addition, some studies have simultaneously analyzed intraspecific and interspecific variation in cognitive performance. In the following subsections, we first give an overview of these studies. We refer to general factors extracted from intraspecific studies as *g*, and to those extracted from interspecific studies as *G*. We then critically assess to what extent alternative explanations may account for the findings, and formulate criteria for future studies that should help pin down to what extent a statistically derived *g*/*G* factor reflects general intelligence as broadly defined.

2.1. Intraspecific studies of psychometric intelligence: *g*

Interest in the question of whether general intelligence may be found in nonhuman animals briefly spiked in the 1930s and 1940s (Locurto & Scanlon 1998), after Spearman's *g* factor (Spearman 1927) had become widely known. These studies reported positive correlations across various types of tasks, but predominantly concerned mazes and mostly in non-primate species such as mice, rats, and chicks (Locurto 1997). Because the model of a hierarchical structure of human cognition and the methodological tools to detect it became widely available only in the late 1940s, the design of these early studies was often not suitable to detect *g* or any factor structure.

For the next half century, the question of animal general intelligence was largely ignored, with interest resurging only after the late 1990s, mainly focusing on mice and primates. Table 3 provides an overview of these studies that have assessed and analyzed correlated performance across three or more cognitive tasks within subjects of the same species, for rodents, primates, and other species (see also Bouchard 2014; Chabris 2007; Galsworthy et al. 2014; Matzel et al. 2013).

In rodents, robust evidence for *g* is available from a range of studies, mostly on mice, from test batteries including as many as eight different tasks and various regimes of principal component analysis (e.g., reviewed in Bouchard 2014; Galsworthy et al. 2014; Matzel et al. 2011b; but see Locurto et al. 2003; 2006). In general, *g* explains between 30% and 40% of variation in cognitive performance, and in rats, it is positively correlated with brain size (Anderson 1993). Moreover, heritability estimates of up to 40% have been reported (Galsworthy et al. 2005). Test batteries often include typical, rather basic learning tasks, such as associative fear conditioning, operant avoidance, path integration, odor discrimination, and spatial navigation. Nevertheless, as in humans, the derived *g* factors have been shown to covary with executive functions, such as selective attention (Kolata et al. 2007; Matzel et al. 2011a) and working memory (particularly working memory capacity: Kolata et al. 2005; Matzel et al. 2008; Sauce et al. 2014) as well as performance in tests of reasoning. For instance, *g* derived from a standard mouse test battery predicted performance in inductive (finding efficient search strategies in a complex maze) and deductive reasoning (inferring the meaning of a novel item by exclusion, i.e., "fast mapping": Wass et al. 2012). Working memory training did increase *g* (Light et al. 2010; Matzel et al. 2011a), mainly through its positive effect on selective attention (Light et al. 2010; see also Sauce et al. 2014). Importantly, *g* did not simply capture fear and stress reactivity (Matzel et al. 2006), anxiety (Galsworthy et al. 2002), or other lower-level biological processes such as sensory or motor abilities (Matzel et al. 2006). In sum, for rodents, the finding of a first component in cognitive test batteries that corresponds to *g* is robust, and several implications of its presence have been confirmed.

In nonhuman primates, only a handful of studies on the consistency of individual-level differences in cognitive tasks are available. Herndon et al. (1997) were interested in classifying patterns of age-related cognitive decline in adult rhesus macaques, an Old World monkey species. They found a first PCA factor that explained 48% of the variance in cognitive performance and on which all six tasks loaded

Table 3. *Intraspecific studies that have assessed and analyzed correlated performance across at least three cognitive tasks within subjects of the same species, for rodents, primates, and other species*

	Species (*n*)	Test Battery	Key Findings and Conclusion	Reference
Rodents	Rats (22 + 20[1])	4 tasks: attention to novelty, speed, and accuracy of reasoning (8-arm radial maze), response flexibility (detour problem)	Evidence for *g* in both samples; *g* was correlated with brain weight (second sample).	Anderson (1993)
	Mice (two strains: 34 + 41)	5 water escape tasks: route learning (Hebb-Williams maze), use of spatial navigational cues (Morris water maze), spatial reversal learning and visual reversal learning (T-maze), place learning (4-arm maze); *plus* activity control task	Evidence for *g* in both strains (explaining 61% and 55% of variance in the latency measures, and 28% and 37% in the error measures); authors stress limited implication for *g* because mainly spatial tasks were used; activity loads on first factor in strain A but not in strain B.	Locurto and Scanlon (1998)
	Mice (40)	6 tasks: curiosity (spontaneous alternation in T-maze), route learning (Hebb-Williams maze), use of spatial navigational cues (Morris water maze), detour problem (burrowing task), contextual memory, plug puzzle; *plus* anxiety in new environments (open field)	Evidence for *g* (explaining 31% of variance); *g* was independent of anxiety.	Galsworthy et al. (2002)
	Mice (60)	6 tasks: route learning (Hebb-Williams), place learning (plus maze), and a set of detour problems; 3 working memory tasks (8-arm radial maze, 4 × 4 radial maze, visual non-matching to sample), *plus* 3 activity and stress control tasks	No evidence for *g* (first factor explains 19.4% of variance, control tasks included in PCA).	Locurto et al. (2003)
	Mice (56)	Standard mouse battery of 5 tasks: associative fear conditioning, operant avoidance, path integration (Lashley III maze), odor discrimination, and spatial navigation (spatial water maze) *plus* open field exploration task	Evidence for *g* (explaining 38% of variance); exploration propensity related to individual learning ability.	Matzel et al. (2003)
	Mice (21)	Variant of standard mouse battery *plus* exploration task (open field), long-term retention (retest in Lashley III maze after 30 days) and working memory task (simultaneous performance in two 8-arm radial mazes)	Evidence for *g* (explaining 43% of variance); *g* covaried with exploration and working memory capacity but not with long-term retention.	Kolata et al. (2005)
	Mice (84 unrelated,[1] and 167 siblings)	Tasks from Galsworthy et al. (2002) *plus* object exploration and 2nd problem-solving task	Evidence for *g* (explaining 23%–41% of variance); *g* showed sibling correlations of 0.17–0.21 and an estimated heritability of 40% (upper limit).	Galsworthy et al. (2005)
	Mice (47 + 51)	Exp. 1: 5 tasks: detour, win-shift, olfactory discrimination, fear conditioning, and operant acquisition; *plus* open field and light-dark control tasks Exp. 2: similar but optimized task battery (same detour and fear conditioning but 3 new tasks, including working memory); same control tasks	Evidence for *g* (explaining 28%–34% of variance) but only after removing control procedures from the analysis; *g* was stronger in the second experiment.	Locurto et al. (2006)
	Mice (43)	Standard mouse battery; *plus* 21 tests of exploratory behavior, sensory/motor function (e.g., running and swimming speed, balance tasks, grip strength) and fitness, emotionality, and hormonal and behavioral stress reactivity	Evidence for *g* (explaining 32% of variance); open field exploration and 7 other explorative behaviors also loaded on this first factor, but *g* was not correlated with general activity, sensory/motor function, physical characteristics, or direct measures of fear; lower-level biological properties loaded weakly and inconsistently on *g*.	Matzel et al. (2006)

(*continued*)

Table 3 (*Continued*)

Species (*n*)	Test Battery	Key Findings and Conclusion	Reference
Mice (27)	Standard mouse battery; *plus* selective attention (complex discrimination), short-term memory capacity (nonspatial radial arm maze), short-term memory duration (delayed reinforced alternation)	Evidence for *g* (explaining 44% of variance); *g* was most strongly correlated with selective attention, followed by simple memory capacity and only weakly with short-term memory duration.	Kolata et al. (2007)
Balb/C Mice (56)	Standard mouse battery *plus* working memory span and capacity, and 12 non-cognitive tests of unlearned behaviors and fitness	Evidence for *g* (explaining 31% of variance); old subjects (19–21 months of age) had lower *g* than young ones (3–5 months of age) but also showed higher variability. Working memory capacity and duration explained variance in *g*, and particularly so in old mice. Old mice with age-related cognitive decline had increased body weight and decreased activity. Some non-cognitive variables were also correlated with *g*.	Matzel et al. (2008)
Mice (69)	Standard mouse battery as adults; *plus* extensive exposure to 12 novel environments prior to testing	Evidence for *g* (explaining 27% of variance); exposure to novelty as juveniles (from 39 days of age) and young adults (from 61 days of age) increased exploration but did not affect *g* compared to control groups when tested as adults (from 79 days of age).	Light et al. (2008)
Mice (241)	Standard mouse battery; subsample of 78 subjects also tested with 2 additional spatial tasks (win-stay and reinforced alternation)	Evidence for *g* (explaining 38% of variation); identification of an additional domain-specific factor for tasks that depended on hippocampal/spatial processing in subsample.	Kolata et al. (2008)
Mice (60)	Standard mouse battery; *plus* prefrontal cortex gene expression profiles	Evidence for *g* (explaining 41%–42% of variance); dopaminergic genes plus one vascular gene significantly correlated with *g*; D1-mediated dopamine signaling in the prefrontal cortex was predictive of *g*, arguably through its modulation of working memory.	Kolata et al. (2010)
Mice (29)	Standard mouse battery; *plus* extensive training on short-term memory duration and working memory capacity, and a selective attention task (Mouse-Stroop)	Evidence for *g* (explaining 30% of variance); working memory training promoted *g*, largely but not exclusively via increased selective attention; effects were smaller when selective attention load of training task was reduced.	Light et al. (2010)
Mice (42)	Standard mouse battery: *plus* 2 exploration tasks (open field and novel environments)	Evidence for *g* (explaining 40% of variance); link between *g* and exploration propensity was mediated by different rates of habituation in high vs. low *g* subjects.	Light et al. (2011), experiment 2
Mice (26)	5 tasks: acquisition of three learning tasks (passive avoidance, shuttle avoidance, reinforced alternation), reversal learning, and selective attention; *plus* longitudinal working memory training (radial arm maze task with overlapping cues, various regimes) and four non-cognitive variables	Evidence for *g* (explaining 26%–37% of variance); longitudinal working memory training prevented age-related decline of attention, learning abilities, and cognitive flexibility; non-cognitive variables loaded moderately to weakly on *g* and in a non-consistent manner; old (from 18 months of age); young (from 5 months of age).	Matzel et al. (2011a; 2011b)

	Mice (47)	Standard mouse battery; *plus* deductive reasoning (inferring by exclusion: fast mapping) and inductive reasoning (efficient search strategy)	Evidence for *g* (explaining 27%–32% of variance); *g* correlated with inductive and deductive reasoning performance.	Wass et al. (2012)
	Mice (26)	4 learning tasks: odor discrimination, reinforced alternation, fear conditioning, radial arm maze *plus* attention battery consisting of 4 tasks: Mouse-Stroop (conflicting visual and olfactory cues), T-maze reversal, coupled latent inhibition, and dual radial arm maze	Evidence for *g* (explaining 37% of variance); different types of attention (external: selective attention; internal: inhibition) contributed independently to variation in *g*.	Sauce et al. (2014)
Primates	Rhesus macaques (30+23)	6 non-social tasks (n = 30): delayed non-matching to sample (acquisition time and performance after 120 sec delay), delayed recognition span task (spatial and color condition), and reversal learning task (spatial and object condition)	Evidence for *g* (explaining 48% of variance), *g* but none of the other two extracted factors declined with age.	Herndon et al. (1997)
			Age groups (age in years): young adults (<15), early-aged (19–23), advanced aged (24–28), and oldest aged (≥29).	
		Subset of the 6 tasks above (n = 53): acquisition and 120″ performance in delayed non-matching to sample, spatial delayed recognition span	Evidence for *g* (explaining 62% of variance); *g* declined with age and was strongly correlated with *g* extracted from the full test battery.	
	Cotton-top tamarins (22)	11 mostly non-social tasks[3]: 10 from the physical domain, 1 from the social domain	Evidence for *g* (Bayesian latent variable approach) but no additional group factors (domains).	Banerjee et al. (2009)
	Chimpanzees (106), 2-year old children (105)	15 of the 16 tasks of the PCTB[4] from the physical and social domain (tool use excluded)	Confirmatory factor analysis revealed different factor structures for chimpanzees (factor 1: spatial tasks; factor 2: some physical and some socio-cognitive tasks) and children (factor 1: spatial tasks; factor 2: some physical tasks; factor 3: 6 social tasks);	Herrmann et al. (2010b)
			Inconclusive regarding *g* for both human children and chimpanzees because of inclusion of social domain and low variability in performance in some of the tasks.	
	Chimpanzees (99)	13 of the 16 tasks of the PCTB[4] from the physical and social domain (without the number addition, social learning, and intention task)	Evidence for *g* (Parallel analysis); *g* was heritable (heritability h^2 = 0.525, p = 0.008).	Hopkins et al. (2014)
			Individual differences in cognitive performance and heritability remained stable in a retest after two years (n = 86).	
			Evidence for *g* (loadings of tasks on first factor range from 0.048–0.607). Subtests with higher *g* loadings were more heritable, and performance in these subtests was more variable between individuals.	Woodley of Menie et al. (2015)
Other species	Dogs (13)	3 tasks: response latencies in discrimination, reversal learning, and visuo-spatial memory (3 delayed non-matching to sample conditions)	Highly significant correlations of performance across all 3 tasks.	Nippak and Milgram (2005)
	Dogs (68 border collies)	6 tasks: four detour tasks, human point following, and numerical discrimination	Evidence for *g*; confirmatory factor analysis on 8 variables (4 detour performance plus speed and choice in point and discrimination task), with latent factors navigation speed, choice speed, and choice accuracy, best fit for hierarchical model with *g* explaining 17% of variation.	Arden and Adams (2016)

(*continued*)

Table 3 (*Continued*)

Species (*n*)	Test Battery	Key Findings and Conclusion	Reference
Bowerbirds (21)	6 ecologically relevant tasks (2 problem solving, 1 mimetic repertoire, and 3 bower building tasks)	Weak evidence for *g* (explaining 27.5% of variance), but *g* as well as separate performance in 4 tasks was correlated with mating success.	Keagy et al. (2011)
Bowerbirds (11)	6 cognitive tasks (an ecologically relevant bower maintenance task, 2 discrimination tasks, and a reversal learning, a spatial memory, and a novel motor task)	Evidence for *g* (explaining 44% of variance), but no correlation between *g* or separate performance in any task with mating success.	Isden et al. (2013)
New Zealand Robins (16)	6 ecologically relevant tasks (1 motor task, color and shape discrimination, reversal learning, spatial memory, and inhibitory control); *plus* motivation (sitting on an electronic scale and eating a mealworm) and neophobia (latency to touch the new apparatus)	Evidence for *g* (explaining 34% of variance). Consistent pattern of results after removing the spatial memory task or/and subjects with a color preference. Even stronger evidence for *g* (45%) when removing both the motor task and subjects with a color preference from the PCA.	Shaw et al. (2015)

[1] In the second sample, variation in brain size was induced by prenatal exposure to methyl-azoxymethanol, which induces microcephaly
[2] Data for 40 subjects were taken from Galsworthy et al. (2002)
[3] Inhibition (occluded reach, A-not-B error, reversal learning), perceptual speed (targeted reach), exploration, numerical discrimination, acoustic discrimination, inspection time (objects and social), memory (hidden reward retrieval), food extraction puzzle
[4] The PCTB (Primate Cognition Test Battery, Herrmann et al. 2007) consists of 16 tasks from the physical domain (space: spatial memory, object permanence, rotation, transposition; quantities: relative numbers, addition numbers; causality: noise, shape, tool use, tool properties) and the social domain (social learning; communication: comprehension, pointing cups, attentional state; theory of mind: gaze following, intentions)

positively. This factor, based on 30 subjects, was highly correlated with a factor derived from a subset of only three of these tasks (all of which, again, loaded positively on it) in an overlapping sample of 53 subjects. Furthermore, this putative *g* declined linearly with increasing age of the monkeys.

Banerjee et al. (2009) found evidence for *g* in a New World monkey species, the cotton-top tamarin (*Saguinus oedipus*). They tested 22 subjects with a battery consisting of 11 tasks that assess a range of cognitive abilities such as inhibitory control, quantity discrimination, and memory. Owing to the relatively high number of missing individual test scores, they used Bayesian analysis and found a *g* factor but no group factors that would have corresponded to more specialized cognitive domains (although the a priori classification of domains is inevitably tenuous without extensive validation; see also section 2.4).

Among great apes, evidence for *g* is more mixed. Herrmann et al. (2007) developed the Primate Cognitive Test Battery (PCTB) consisting of 16 tasks from the physical and the social domain, a priori placed into six categories (i.e., space, quantities, causality, social learning, communication, and theory of mind) and applied it to 106 chimpanzees, 32 orangutans, and 105 two-year old human children. Chimpanzees and human children performed equally well (and better than orangutans) in tasks from the physical domain, but the children outperformed both ape species in the social domain. These results were not consistent with *g* in any of the species, including human children. To explicitly address the structure of individual differences, Herrmann et al. (2010b) re-analyzed the data from the chimpanzees and children in 15 of the 16 PCTB tasks (tool use was not included) using a confirmatory PCA (see sect. 2.4.1 for further discussion). They found a different structure of cognitive abilities for chimpanzees (2 factors) and children (3 factors). In addition to a "Spatial" factor in both species, only one additional "Physical-Social" factor emerged in chimpanzees, whereas two additional factors, a "Physical" and a "Social" one, emerged in children. The authors thus did not find evidence for *g* in either chimpanzees or humans. However, human test batteries typically do not include subtests assessing social cognition. In fact, the relationship between general cognitive processes and socio-cognitive processes is currently poorly understood in humans (Korman et al. 2015). This problem, however, does not explain the presence of two other factors rather than a single *g* in human children.

More recently, Hopkins et al. (2014) tested 99 chimpanzees with a reduced and slightly modified version of the PCTB consisting of 13 of the 16 tasks (including tool use but excluding one of two quantity tasks, the social learning task, and one theory of mind task). They report a *g* factor derived from a non-rotated PCA and used quantitative genetic analyses to estimate its heritability (h^2), which was found to be 53% and highly significant. Furthermore, the results remained stable when 86 of the 99 chimpanzees were retested with the same test battery after two years, and were confirmed with parallel analysis. Woodley of Menie et al. (2015) further analyzed the data set and concluded that the more *g*-loaded a task is, the higher its heritability and phenotypical variability, as also found in humans. The more *g*-loaded tasks also had higher coefficients of additive genetic variance, suggesting that cognitive abilities with higher *g* loadings have been subject to stronger recent selection.

Taken together, then, the psychometric studies in specific rodent and primate species lend increasing support to the notion that the positive manifold is not unique to humans but also present in nonhuman animals. Studies on other lineages such as dogs (Arden & Adams 2016; Nippak & Milgram 2005) and birds (Keagy et al. 2011; Isden et al. 2013; Shaw et al. 2015) are also beginning to provide evidence. However, a serious limitation of psychometric studies in nonhuman animals is that they tend to lack power with respect to sample size, the diversity of cognitive tasks, or both. We discuss these limitations in section 2.4. Fortunately, there is a complementary approach, which examines interspecific variation and is particularly powerful to reveal evolutionary trends.

2.2. Interspecific studies of psychometric intelligence: G

In comparative approaches, the fundamental question is whether some species systematically outperform others across an array of distinct cognitive tasks, consistent with the notion of psychometric and perhaps general intelligence, or whether species differences are instead characterized by independent variation in performance across tasks and domains, consistent with higher domain specificity. Comparative studies thus investigate whether what evolves are specialized skills or rather general intelligence. This approach (Table 4) has predominantly been applied to primates but also to birds and involves both meta-analyses and targeted experimental comparisons.

For primates, Deaner et al. (2006) conducted a meta-analytical study that compared the performance of 24 primate taxa tested with nine experimental physical-cognition paradigms using Bayesian hierarchical modeling (Johnson et al. 2002). They found strong evidence for *G*, which correctly predicted 85% of the species rankings (but note that caution is needed when comparing the proportion of explained variance between standard PCA and Bayesian analyses). Moreover, in a follow-up study, *G* was strongly correlated with brain size (Deaner et al. 2007).

In another set of studies, Reader and Laland (2002) collected data from the literature on the incidence of innovation, social learning, and tool use in 116 species of

Table 4. *Interspecific, comparative studies that have assessed correlated cognitive performance across species*

Species (*n*)	Type of Study	Key Finding	Reference
Primate species (116)	Correlation of ecologically relevant cognitive abilities (innovation, tool use and social learning) and volume measures of the executive brain (neocortex and striatum) and brainstem (mesencephalon and medulla oblongata)	The 3 measures were correlated across nonhuman primate species and with both absolute and relative executive brain volumes; results consistent with *G*	Reader and Laland (2002)
Primate taxa (24) (3 great ape species, 1 lesser ape, and 7 catarrhine, 6 platyrrhine, & 7 prosimian genera)	Meta-analysis of 9 experimental paradigms (detour problems, patterned-string problems, invisible displacement, tool use, object discrimination learning set, reversal learning, oddity learning, sorting, and delayed response) of captive subjects using hierarchical Bayesian latent variable analysis (Johnson et al. 2002)	Species- *G* explained 85% of variance; great apes (*Gorilla*, *Pan*, *Pongo*) outperformed all other genera; *G* was positively correlated with various measures of brain size	Deaner et al. (2006; 2007)
Primate species (62) (including apes, catarrhine and platyrrhine monkeys, & prosimians)	Meta-analysis on ecologically relevant tasks: behavioral innovation, social learning, tool use, extractive foraging (expanded data set from Reader & Laland 2002), and tactical deception (data from Byrne & Whiten 1990) using principal component, factor, and phylogenetic analyses	Species- *G* explained 65% of the variance in cognitive performance and covaried with brain size. *G* also covaried with results from captive subjects: that is, the species- *G* from Deaner et al. 2006 and learning performance from Riddell & Corl 1977)	Reader et al. (2011)
Primate species (69) (including apes, catarrhine and platyrrhine monkeys, & prosimians)	Meta-analysis of data sets from Reader et al. (2011, innovation, tool use, social learning, and extractive foraging) and Byrne and Whiten (1990, tactical deception) using principal axis factor analysis and unit weighted factor analysis	Differences in cognitive abilities among primates were concentrated on *G* (explaining almost 62% of variance), and this effect was particularly pronounced in catarrhines (i.e., apes and Old World monkeys)	Fernandes et al. (2014)

nonhuman primates, both in captivity and in the wild, and found that across species, all three measures were correlated with each other (r^2 values around 0.4), as well as with brain size. In a follow-up study involving 62 primate species (Reader et al. 2011), they found evidence for general intelligence on the interspecific level (*G*) in principal component and factor analyses explaining 65% of the variance, based on measures of innovation, social learning, and tool use, as well as extractive foraging and tactical deception. As in Reader and Laland's (2002) earlier study, *G* was correlated with brain size, but also with a combined measure of performance across several learning tasks, with learning set performance (both taken from Riddell & Corl 1977), and the *G* measure of Deaner et al. (2006).

More recently, Fernandes et al. (2014) compiled published data from five cognitive domains (innovation, tool use, social learning, extractive foraging, and tactical deception) across 62 primate species (data sets from Reader & Laland 2002; and Byrne & Whiten 1990). Fernandes et al. found that a single factor *G* explained almost 62% of the total variance. Furthermore, they reported that cognitive abilities that load more strongly on *G* show bigger interspecific variation, weaker phylogenetic signals, and faster rates of evolution. These results are consistent with the idea that *G* has been subjected to selection pressure stronger than narrow, more domain-specific abilities and that *G* is thus the principal locus of selection in the evolution of primate intelligence (but see also section 2.5).

The only other taxon to which comparative approaches have been applied are birds. As in primates, significant positive correlations across species were found between innovation rates, tool use, and learning performance. These studies also found positive correlations between innovation rates and brain size as well as colonization success (Ducatez et al. 2015; Lefebvre 2013; Lefebvre et al. 2004; Sol et al. 2005).

2.3. Mixed studies combining intraspecific and interspecific variation

Some studies have pursued a mixed approach by applying test batteries to multiple individuals from several species (see Table 5). For instance, Herrmann and Call (2012) analyzed data of 23 individuals from all four nonhuman great ape species, which were studied in a range of tasks from the physical domain, and found no support for the existence of *g*. Nevertheless, some subjects performed particularly well (or poorly) across tasks, both in the sample of 23 great apes and in the 106 chimpanzees mentioned previously (Herrmann et al. 2010b), indicating that there was some consistency in individual performance.

In another mixed study, Amici et al. (2012) found no evidence for *G* or *g* when re-analyzing data from seven primate species (all four great ape species, long-tailed macaques, spider monkeys, and capuchin monkeys, totaling 99 individuals) from 17 cognitive tasks. In the Bayesian approach used to analyze the data (see also Barney et al. 2015), the 17 tasks were a priori attributed to the domains of inhibition, memory, transposition, and support, similar to Herrmann et al. (2010b).

In contrast to intraspecific and interspecific studies, mixed studies thus provide less support for psychometric intelligence. Here, we offer a tentative suggestion to

Table 5. *Mixed studies that have simultaneously analyzed correlated performance within and between species*

Species (*n*)	Type of Study	Key Finding	Reference
Chimpanzees (106) Orangutans (32) 2.5-year-old human children (105)	Psychometric study using the Primate Cognitive Test Battery (PCTB) consisting of 16 tasks from the physical domain (space: spatial memory, object permanence, rotation, transposition; quantities: relative numbers, addition numbers; causality: noise, shape, tool use, tool properties) and the social domain (social learning; communication: comprehension, pointing cups, attentional state; theory of mind: gaze following, intentions) using analysis of variance	Chimpanzees and human children performed equally well (and better than orangutans) in the physical domain, but the children outperformed both ape species in the social domain; results not consistent with *G*	Herrmann et al. (2007)
Bonobos, chimpanzees, gorillas, and orangutans (23)	8 non-social tasks from various studies: spatial knowledge (i.e., delayed response, inhibition, A-not-B, rotations, transpositions and object permanence), tool use (4 tests), inferential reasoning by exclusion, quantity discrimination, causal reasoning and color, size and shape discrimination learning	No evidence for *g*; but some individuals performed consistently well across tasks	Herrmann and Call (2012)
Chimpanzees (19), orangutans (10), bonobos (5), gorillas (8), long-tailed macaques (12), spider monkeys (18), capuchin monkeys (27)	Re-analysis of data obtained from two psychometric studies resulting in 17 tasks from four physical domains (inhibition from Amici et al. [2008; 2010], and memory, transposition, and support from Herrmann et al. [2007]) with captive subjects using a hierarchical Bayesian modeling approach	Most variance explained by species and cognitive domain; results not consistent with *G*	Amici et al. (2012); Barney et al. (2015)

explain this absence of evidence for psychometric intelligence in mixed studies that will need to be examined in more detail in future work. First, despite including a large number of individuals overall, the effective sample size to identify *g* remains the number of individuals within each species, and to identify *G* is the number of species. This may strongly influence the outcome because in mixed studies the detection of *G* is not based on average species-specific performance as is done in interspecific studies, but is instead based on individual values, which are more susceptible to noise. A recent memory task illustrates the superiority of species averages in estimating abilities at the species level. In this study, both marmoset and squirrel monkeys as a group provided results fitting the Ebbinghaus forgetting curve, but at the individual level, several individuals did not, indicating that the performance of these individuals was strongly affected by noise (Schubiger et al. 2016). Such noise may overshadow *G*, especially in species that are very close in *G*.

A second issue is that in a sample of species with similar *G* (e.g., according to Deaner et al. 2006), and thus both the great ape study by Herrmann and Call (2012) and the study by Amici et al. (2012), species-specific predispositions linked to domain-specific adaptations may mask a *G* effect. Thus, chimpanzees and orangutans are more extraction-oriented than bonobos or gorillas (van Schaik 2016), as expressed, for instance, in species differences in tendencies to handle objects (Koops et al. 2015), or to solve social problems (Herrmann et al. 2010a). Such variation is bound to produce species differences in mean performance on some but not on other tasks, reducing the correlation across tasks in the overall data set. Intraspecific comparisons obviously are not affected by this problem, whereas the effect on interspecific comparisons is reduced the broader the comparison in terms of *G* are, because major interspecific differences in *G* lessen the effects of species differences in domain-specific predispositions.

More generally, we can ask, if in a given lineage, a robust *G* is found, whether this implies that all species in the comparison *must* have *g*, and vice versa. Several combinations of evidence for *g* and *G* are possible, in particular in mixed studies, as summarized in Table 6, and we discuss likely explanations for these combinations.

The interpretation is straightforward whenever evidence for *g* and *G* point in the same direction (see entries I and IV in Table 6), and where it is positive for both, can be externally validated separately at both the level of *g* and *G* (see also section 2.5). One potentially conflicting constellation is when positive evidence for *g* but no evidence for *G* is available (as for entry II in Table 6). Such a result can arise if *g* is present in only a few of the species involved in the comparison, which might occur when distantly related lineages are compared. The other conflicting constellation (entry III in Table 6) is that comparative studies provide evidence for *G*, but there is no evidence for *g* within the species involved in the comparison. This was the case in some primate studies. In principle, it is possible that we are dealing with cumulative modularity and that by chance the distribution of modules across the species included in the sample is hierarchically nested. In this scenario, no correlation between *G* and EFs or, arguably, brain size is expected, which is inconsistent with current findings. The most likely cause of constellation III, therefore, is lack of power of animal studies to reliably detect the absence of *g*, due to the small sample sizes and difficulties to construct a suitable test battery, which make animal psychometric *g* studies prone to Type II errors (see also section 2.4.1).

Taken together, there is increasing evidence for *g* in nonhuman animals, particularly in mice and primates, for which positive evidence is available for New World monkeys, Old World monkeys, and chimpanzees (but see Herrmann et al. 2010b). At the interspecific level based on comparative analyses across species, studies of primates and birds provide a robust pattern consistent with *G*. Finally, mixed studies in primates that simultaneously analyze within- and between-species variation yield a more ambiguous pattern.

2.4. Facts or artifacts?

A legitimate concern is whether a presumptive *g*/*G* factor can arise as an artifact, and a legitimate question is to what, exactly, it corresponds. We now review why statistical or methodological artifacts may produce false positives, whereas secondary modularization may lead to false negatives, and formulate criteria for future directions that may be used to evaluate whether g/G corresponds to general intelligence broadly defined.

2.4.1. Statistical issues. The use of PCAs or related procedures involves a suite of decisions, including whether exploratory or confirmatory analyses are applied, whether non-rotated or rotated factors are considered, and whether oblique or orthogonal rotations are used. A

Table 6. *Summary of the potential combinations of evidence for* g *and* G*, and under what conditions apparently conflicting findings can be reconciled*

	No Evidence for *g*	Evidence for *g*
No evidence for *G*	**I**: domain-specific cognitive abilities	**II**: *g* is present in only a few of the species involved in the comparative approach; or the involved species are very close in *G* and evidence for it is masked by variation in species-specific predispositions
Evidence for *G*	**III**: largely cumulative modularity; or artifact due to lack of power of animal psychometric studies	**IV**: general intelligence, in particular if supported by external validation of both *g* and *G*

detailed discussion of factor-analytical procedures is far beyond the scope of this review, and we refer readers to the specialized literature (e.g., Barney et al. 2015; Garson 2013; Stevens 2012). However, because these decisions may critically affect the conclusions of animal studies, we must highlight some issues that appear relevant to the empirical results summarized previously.

First, the use of confirmatory analyses requires an a priori decision of what a domain is, and which tasks are associated with the respective domains (this also applies to Bayesian approaches that likewise categorize tasks a priori to hypothesized domains: Amici et al. 2012; Barney et al. 2015). The identification of domains of animal cognition, however, is not straightforward. For instance, some classify spatial reversal learning tasks as spatial cognition (e.g., Locurto & Scanlon 1998) whereas others stress their inhibition component (Tapp et al. 2003). In reality, of course, subjects may recruit several specific abilities to solve a particular task, and in fact different subjects may even recruit a different mix. Accordingly, Hopkins et al. (2014) found that their exploratory PCA findings were not entirely consistent with the a priori structure of the PCTB originally proposed by Herrmann et al. (2007; 2010b). An a priori allocation of tasks to domains is thus not straightforward; in fact, the structure of a species' cognition is an empirical question (see also sect. 1.1.1 for corresponding efforts in human intelligence research). Accordingly, the use of confirmatory techniques may lead to diverging results compared to analytical approaches that are a priori agnostic with regard to factor structure.

Second, studies vary with regard to whether they present rotated or non-rotated solutions. Because rotations are designed to make the pattern of factor loadings more pronounced, it is generally recommended to use non-rotated solutions in *g* studies (Galsworthy et al. 2014; Jensen & Weng 1994; Locurto et al. 2003; Plomin 2001; Woodley of Menie et al. 2015). Rotated and non-rotated solutions from the same data set are presented in Hopkins et al. (2014) and Woodley of Menie et al. (2015). Whereas the varimax-rotated solution (Hopkins et al. 2014, Table 1) appeared to suggest that a general factor *g* was lacking, the results of non-rotated solutions, verified by parallel analysis, demonstrated it was in fact present.

Third, a common intuition in general intelligence studies on animals is to compare the amount of variance explained by a first factor, and to conclude that the higher the amount of explained variance, the stronger the evidence for *g*. In human studies, the first non-rotated factor typically accounts for about 40% of variance (Plomin 2001), which is in fact similar to what has been reported for mice (see Table 3). However, an exclusive focus on the amount of explained variance is problematic for empirical and conceptual reasons. Empirically, the proportion of explained variance not only depends on the statistical issues discussed previously, but also on the heterogeneity of the subjects in the sample: the more heterogeneous, the higher the proportion of variance explained. In interspecific investigations, for instance, this means that studies that involve species that vary widely in general intelligence and brain size (e.g., 20 species of primates ranging from great apes to prosimians) will find higher proportions of explained variance than studies with a similar sample size, but where the species are all relatively similar (e.g., 20 different species from the same genus or taxonomic family). Conceptually, to the extent that the mind is a combination of both specialized cognitive adaptations and domain-general processes (see also sect. 4.1), very small proportions of explained variance may still be indicative of a real *g*. Likewise, a first factor with high loadings of some tasks but not others may reflect the absence of general intelligence, but may also reflect the co-occurrence of a general factor and one or several additional, more specialized domains (e.g., for spatial orientation, see Herrmann et al. 2007; see also first PCA factor in Hopkins et al. 2014).

Last but not least, the most severe statistical restriction of nonhuman psychometric studies is that they critically lack power due to their small sample sizes. Reaching a near-consensus about the structure of human intelligence required meta-analyses involving thousands of subjects (Carroll 1993). Obtaining sample sizes comparable to human studies is unrealistic for most nonhuman animal species, in particular for nonhuman primates (albeit less so for rodents). However, replicating studies is feasible, and if this reveals the same factorial solution in a different set of subjects, and if combining such data sets also increases the fit of the solution, we can be increasingly confident that we are not dealing with statistical artifacts. Unfortunately, although this approach minimizes Type I errors, it suffers from very limited power to avoid Type II errors. In other words, if successful, we can be confident that we have obtained a real result, but if it fails, this may reflect either the absence of a general factor or too low a number of subjects. This shortcoming highlights the need to use external validation for psychometric *g*/*G* studies, as discussed below in section 2.5.

2.4.2. Methodological issues. We now turn to the possibility that a *g*/*G* factor may arise as a methodological artifact, because the results reflect variation in underlying variables other than general intelligence (see also Macphail's [1982] contextual variables) or because the tasks mainly tap into problems of the same domain.

Some individuals, or some species, may systematically outperform others not because they are more intelligent, but because they are less fearful and better habituated to testing, are more motivated to participate in tasks, have sharper senses, or are simply more active than others (Macphail 1982). Ideally, such confounds are directly quantified, as for instance in Matzel et al. (2006). In a sample of 43 mice individuals, they examined to what extent the general learning ability *g* extracted via PCA from a test battery of six cognitive tasks was correlated with 21 measures of exploratory behavior, sensory/motor function (e.g., running and swimming speed, balance tasks), activity, or fear/stress sensitivity. They found that *g* was not explained by general activity, sensory/motor function, physical characteristics, or direct measures of fear, but was correlated with several exploratory behaviors. Follow-up studies suggested that this link is caused by variation in habituation rates when exposed to potentially stressful situations (Light et al. 2011) rather than by fearfulness influencing both exploration and task performance: Treatment with anxiolytic drugs did increase exploratory behaviors but did not improve performance in individual tasks or *g* (Grossman et al. 2007). Likewise, temporary environmental enrichment resulting in increased exploration tendency did not improve performance on the cognitive test battery (Light et al. 2008). Thus, exploration and *g* may covary

because more exploratory individuals are more likely to encounter contingencies in the environment that promote learning and problem solving, which over time leads to greater experience. The correlation between exploration and *g* may thus reflect a long-term, cumulative effect of experience on *g*. This is in line with investment theory (Cattell 1987), and with findings in human infants, where the preference for novelty and habituation is positively correlated with later performance in IQ tests (Teubert et al. 2011), but also with apes, where individuals more likely to approach novel objects and a human stranger performed better in physical-cognition tasks (Herrmann et al. 2007). Thus, the rodent studies support the idea that *g* is not an artifact of confounding factors.

Another non-cognitive factor that may explain variation in cognitive performance is motivation to participate. Female callitrichid monkeys have been reported to outperform males in problem-solving tasks (Brown et al. 2010; Yamamoto et al. 2004). However, female callitrichids are typically also more food-motivated, whereas males are more vigilant than females (Koenig 1998). Accordingly, males are less interested in participating in experimental tasks and more easily emotionally aroused during testing. But if male performance is controlled for the presence of attention to the test stimuli, their performance is no longer inferior to that of females (Schubiger et al. 2015). The sexes thus do not differ in cognitive ability, but in their motivation to participate in experimental tasks.

The problem that we may never be sure if species differences in cognitive performance are the result of differences in cognitive ability or differences in contextual variables (Macphail 1982) remains an ongoing challenge for any species comparison. Nevertheless, not all tasks are affected by this problem to the same extent. Reversal learning tasks, for instance, are arguably less affected, because individuals first have to reach a criterion of making an initial discrimination. Differences in sensory-motor abilities and so on may well influence how difficult it is for a species to learn a particular discrimination. However, the crucial test is applied only once a specific criterion has been reached, and at least in marmosets, the time needed to achieve this criterion does not predict performance in the reversal trials (Strasser & Burkart 2012). Furthermore, it is reassuring that the strongest association between a specific task and *G* in Deaner et al. (2006) was the one between reversal learning and *G*.

A second fundamental methodological issue refers to the task selection and battery development. With respect to task design, it is increasingly recognized that small differences in methodological details can strongly influence task performance, which has to be taken into account when performing species comparisons. For instance, memory performance strongly depends on task format in both marmosets and squirrel monkeys. Tests of memory often rely on a two-option choice task (e.g., Banerjee et al. 2009), but many individuals are then happy to follow a random choice, which yields a 50% reward rate. When the choice involves many more options, subjects will be more motivated to remember the location of the food items and provide more-accurate estimates of their ability to memorize the location of the food item (Schubiger et al. 2016). Regarding battery design, if all tasks in the test battery are drawn from the same domain (i.e., a lower-order group factor), rather than from a variety of domains, the positive correlations will reflect a domain-specific ability rather than a more general underlying cognitive factor (*g*/*G*). For instance, a positive manifold across a number of maze tasks is consistent with a spatial factor, but not informative with regard to *g*. The issue of task selection is thus closely linked to the identification of domains in animal cognition, which in fact is part of the empirical question that needs to be addressed in intelligence research in animals in general, by using batteries as diverse as possible and statistical procedures that are a priori agnostic to the underlying factor structure.

2.4.3. False negatives as a result of secondary modularization. Task selection may also bias the result and potentially produce false negatives if tasks prone to secondary modularization are included. Secondary modularization refers to the process that during ontogeny, individuals may specialize on a specific set of problems in a particular domain (Table 2). Problem solving in this domain becomes automatized and thus acquires many features commonly associated with modules rather than domain-general reasoning, particularly fast and frugal information processing, which is independent of reasoning. Thus, despite the presence of *g* in a given species, performance among individuals across domains need not be correlated whenever heterogeneous developmental inputs prevail that lead different individuals to specialize in different tasks (see Fig. 1c). This applies in particular to the small samples typical for nonhuman primate studies.

Prima facie, this situation (Fig. 1c) may seem incompatible with the positive manifold, which is well documented in humans and perhaps other animals. It is important to keep in mind, however, that psychometric studies in humans are typically performed on subject pools with a rather uniform cultural background (the same is also true for the rodent studies performed on lab animals with virtually identical rearing conditions). If, in human studies, the cultural backgrounds of subjects were more diverse (e.g., ranging from Western-industrialized to a variety of hunter-gatherer societies), and only a small number of subjects tested, such an outcome (as in Fig. 1c) is quite likely (see also Reyes-García et al. 2016). The notorious difficulty of devising culture-free or at least culture-fair intelligence tests is a direct consequence of this problem (Saklofske et al. 2014).

The prime example for secondary modularization in nonhuman primates is tool use, which is part of many test batteries typically used with nonhuman primates (e.g., Herrmann et al. 2007; Reader et al. 2011). Nonhuman primates vary considerably with regard to tool use, with great apes typically outperforming monkeys. But differences also occur within a species, both between wild and captive animals and among wild populations. Individuals of the same species show much higher propensities to use tools in captivity compared to their counterparts in the wild (Meulman et al. 2012; van Schaik et al. 1999). Once proficient, individuals show tool use with high degrees of automatization and efficiency. Wild populations, too, vary significantly with regard to their propensity to use tools and solve tool-related problems (e.g., chimpanzees, Gruber et al. 2011; orangutans, van Schaik et al. 2003; or capuchin monkeys, Cardoso 2013), arguably because they have ontogenetically acquired systematically different affordances of sticks or stones, which are perceived as

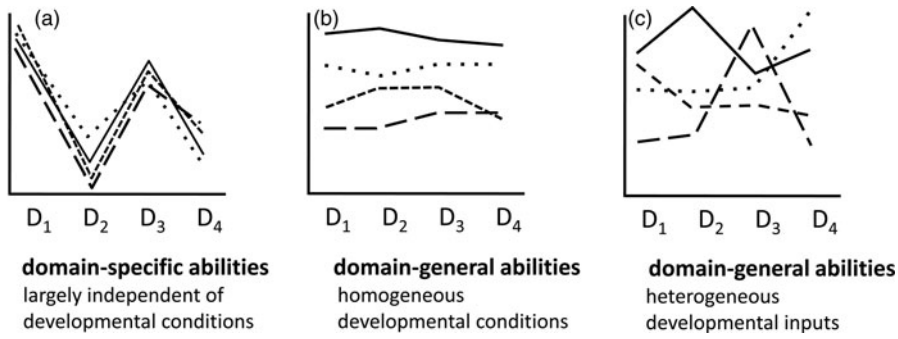


Figure 1. Performance across different cognitive domains (D_1-D_4). Each line represents the performance of an individual (or, in highly cultural species, of a population). (a) Performance is driven by domain-specific abilities; all individuals perform well in some domains but worse in others, but individual differences across domains are random; (b) performance is driven by domain-general abilities and individuals experience homogeneous developmental conditions, which leads to correlated performance between individuals across domains; (c) performance is driven by domain-general abilities but heterogeneous developmental conditions lead to specialization and secondary modularization of individuals in different domains. As a result, performance between individuals across domains is not correlated despite the presence of *g*.

potential tools in habitual tool users but not in non-tool users.

False negatives resulting from secondary modularization (see Fig. 1c) can be minimized if subjects with comparable rearing conditions are selected for the tests, but also if tasks prone to secondary modularization are excluded from test batteries. Thus, instead of naturalistic tasks that test for ontogenetically constructed skills that are likely to become automatized, such as tool use, or the ability to point and understand pointing, or even to use human language systems (Savage-Rumbaugh et al. 2005), it is preferable to include tasks testing for more elementary cognitive abilities, such as reversal learning, mental rotation, or quantity discrimination. Likewise, tests should avoid reliance on experience and knowledge of affordances that may differ among individuals depending on their biographies.

Although it is important to identify tasks and abilities prone to secondary modularization, it is not always easy to identify them. One way to do so is to examine the ontogeny of skills that are suspected to be the result of secondary modularization. Such skills should be acquired by developing immatures after a period of learning (perhaps following alternating series of instances of social learning and practice: Meulman et al. 2013; Schuppli et al. 2016), and could also potentially show high variation among adults. The increasing evidence for a major amount of skill learning by immature primates (e.g., Schuppli et al. 2016) and mammals and birds more generally (van Schaik et al. 2016) suggests a greater prevalence of secondary modularization in nonhumans than revealed by the size of cultural repertoires (Whiten & van Schaik 2007). Because in wild populations, social and ecological problems tend to be very uniform for all individuals, variation of skill profiles (see Fig. 1) between populations (that live under similar, wild or captive, conditions), rather than among individuals of the same population, provides an additional heuristic tool to distinguish between genuine primary and secondary modularity. This criterion would work for primate tool use, for instance. Most powerful to disentangle primary from secondary modularity, finally, are cross-fostering experiments. When cross-fostered individuals exhibit species-typical behavior from the foster species rather than their own species, these behaviors clearly cannot result from primary modules. If the same procedure works within a species at the level of populations, it is similarly evidence for secondary, and thus learned, modules.

2.5. Psychometric or general intelligence? Future directions for animal studies

A crucial question that remains unanswered so far is to what extent a reliably identified *g*/*G* actually captures general intelligence in a broad sense: that is, reasoning ability and behavioral flexibility (Byrne 1994; Gottfredson 1997; Rumbaugh & Washburn 2003; Yoerg 2001; see also sect. 1.1). If it indeed does so, the processes underlying general intelligence (see also sects. 1.1.2 and 1.2.3) in animals should be broadly similar to those found in humans, with the obvious exception of language, and general intelligence should be correlated with independent measures of reasoning ability and behavioral flexibility (see also Bailey et al. 2007). If it is not, the statistically derived psychometric factors may reflect cumulative modularity: that is, the coexistence of separate, but coevolved modules.

These two possibilities can be teased apart empirically: If *g*/*G* represents intelligence in a broad sense, it must be possible to independently assess its validity, for instance, by showing that *g*/*G* is correlated with the domain-general EFs, as has been shown for humans. In principle, an association at a higher hierarchical level (e.g., between-species *G*) may be absent within the subgroups comprising it (e.g., within-species *g*), a phenomenon known as Simpson's paradox (Kievit et al. 2013). In the present case, we may thus find a correlation between *G* and EFs but not between *g* and EFs, which would suggest that *g* and *G* were not aspects of the same phenomenon: that is, general intelligence. Thus, to ensure that *g* and *G* are related to the same phenomenon, one must validate both of them independently.

Intraspecific studies of primates have so far largely neglected the approach to validate *g*, but this approach has provided fruitful insights in rodent studies. In rodents, individual levels of *g* have been shown to correlate with executive functions such as working memory. Matzel

et al. (2008; 2011b) have compared performance on standardized test batteries that reliably quantify *g* in mice with several measures of working memory, including short-term memory duration (how long can the mouse remember which arms of a maze it has already visited?), simple memory span (how many symbols associated with food can the mouse remember?), and selective attention (an adapted version of the STROOP task, in which the subject has to focus on one dimension of the task while suppressing a second dimension that provides conflicting information). As in human studies, they found that *g* was most strongly correlated with selective attention, followed by simple memory span and only weakly with short-term memory duration (Matzel et al. 2008; 2011b; Kolata et al. 2007). Moreover, they showed that training working memory capacity, but not simple working memory span, promotes selective attention and *g* (Light et al. 2010). Future validation tests could also examine the correlation between *g* and conduction speed or the ability to ignore irrelevant, distracting information, which are known correlates of *g* in humans (Melnick et al. 2013; Sheppard & Vernon 2008).

The corresponding validation of psychometrically derived *g*-scores in other species, particularly in nonhuman primates, would be highly desirable. Nonetheless, some evidence consistent with *g* representing domain-general cognitive mechanisms is already available from nonhuman primates. Within chimpanzees, heritability was strongest for overall cognitive performance *g* rather than for distinct aptitudes (Hopkins et al. 2014), as expected when the latter are due to secondary modularization rather than reflect specific adaptations. As a result, cognitive abilities that load higher on *g* in chimpanzees are more heritable, phenotypically variable, and presumably the result of recent natural selection (Woodley of Menie et al. 2015).

The independent evolution of large numbers of modules instead of general intelligence is particularly difficult to reconcile with interspecific findings of *G*. If we are dealing with independent modules, each species would be expected to possess a different repertoire of primary modules, according to the specific adaptive problems it faces. Importantly, across species, this should not result in a stable *G* factor. Studies providing evidence for *G*, however, suggest that particular species generally perform better or more poorly across all tested domains. This is also consistent with the empirical findings suggesting that differences in cognitive abilities among primates are concentrated on *G* (Fernandes et al. 2014). Furthermore, because specific skills, even if complex, can be performed with a very modest amount of brain tissue (e.g., Chittka & Niven 2009; Holekamp et al. 2015; Patton & Braithwaite 2015), one would not necessarily expect that *G* as a reflection of a large number of dedicated modules would correlate with brain size. The well-documented positive correlations between *G* and brain size thus further suggests that *G* reflects general intelligence, as does the finding that across primate species *G* was the principal locus of selection in the macroevolution of intelligence (Fernandes et al. 2014).

Empirical data also support a link between interspecific *G* and executive functions: Across primate species, brain size is not only correlated with *G*, but also with self (inhibitory) control (MacLean et al. 2014). This measure of self-control is directly correlated, in addition (Fig. 2), with *G* as

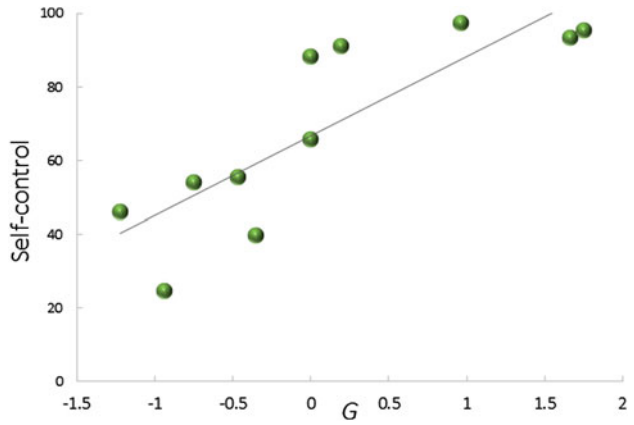


Figure 2. Positive correlation between the composite self-control measure of MacLean et al. (2014) and *G* (Deaner et al. 2006). Adjusted $r^2 = 0.66$, $F_{(1,9)} = 20.75$, $p = 0.001$ based on PGLS (phylogenetic generalized least squares) analyses. The same results are also found when the inhibitory control tasks included in the composite measure are analyzed separately (Cylinder task: adjusted $r^2 = 0.58$, $F_{(1,11)} = 20.75$, $p = 0.002$; A-not-B error: adjusted $r^2 = 0.41$, $F_{(1,12)} = 10.03$, $p = 0.008$).

established by Deaner et al. (2006), which has been derived from a completely independent data set.

A particularly powerful, but likewise underexplored, approach to construct validation consists in training individuals to solve a task in one domain and test to what extent they are able to apply their solution in a different domain. Intraspecific variation in this kind of cognitive flexibility (which is consistent with broad notions of general intelligence as applied by comparative scientists) should be correlated with psychometrically derived measures of individual *g*. Such a pattern would confirm that *g* is indeed a proxy of animal intelligence broadly defined.

Equally promising is to focus on unusually difficult problems relative to individual performance (i.e., problems that cannot be solved in a routine way). For instance, performance in difficult problems such as fast mapping or inductive reasoning was correlated with independently assessed *g* in mice (Wass et al. 2012). Particularly strong evidence would include the demonstration that individuals recruit the same basic cognitive processes for such difficult problems that are also strongly correlated with *g*, such as selective attention or working memory capacity (Matzel et al. 2013; Geary 2009).

Finally, particularly rigorous validation would be based on extra-domain assays. Just as human *g* correlates with academic success, workplace success, health, and even happiness (for references, see section 1), one could in principle ask whether *g* in animals is correlated with outcomes such as the size of cultural repertoires in nature, the ability to rise in social dominance, or to find food during periods of scarcity, and thus survival and even fitness. Unfortunately, this approach is difficult to achieve because it requires both reliably quantifying *g* and the various real-life outcomes in animals under natural conditions. More importantly, *g* may not necessarily predict basic fitness measures such as reproductive success, because of possible tradeoffs between investment into improving general intelligence and other vital activities, such as vigilance or social monitoring.

Table 7 summarizes the issues we have discussed in the form of criteria that may be fruitful to guide future studies.

3. Implications for the evolution of general intelligence

Taken together, although more validation remains to be done, especially in primates, the body of evidence is currently more consistent with the presence of domain-general cognitive abilities in primates and mice, reflected in *g* and *G*, than with the exclusive presence of independent, domain-specific cognitive adaptations. If general intelligence is not limited to humans, this inevitably leads to the question of the conditions favoring the evolution of general intelligence, to which we now turn. Whereas the evolution of Fodorian, specific, dedicated cognitive adaptations in response to recurrent fitness-relevant problems is seemingly straightforward (but see sect. 1.2.1), the evolution of general intelligence poses a puzzle. Domain-specific cognitive adaptations can be instantiated with modest amounts of neural tissue (Chittka & Niven 2009; Holekamp et al. 2015; Patton & Braithwaite 2015) and directly bring about fitness-relevant benefits. Domain-general cognitive ability, however, seems to require substantial amounts of additional expensive brain tissue (Deaner et al. 2007; Reader et al. 2011), and is not automatically linked to fitness-relevant benefits because survival-relevant skills have to be ontogenetically constructed during a process of learning (van Schaik & Burkart 2011). This ontogenetic construction may be more successful in individuals with higher cognitive ability, as posited by investment theory (Cattell 1987), but additional factors also come into play, which renders the link between cognitive ability and fitness-relevant skills more fragile. For instance, whether a survival-relevant skill is acquired may also depend on pure chance (van Schaik & Burkart 2011; van Schaik et al. 2016). Furthermore, in order to more reliably translate general cognitive ability into fitness-relevant skills, some mechanisms for adaptive canalizations (as highlighted in sect. 1.2.2, Table 1) remain necessary, which have to coevolve or, if already present, be linked to the evolving domain-general cognitive processes. We are, therefore, faced with the puzzle that domain-general cognitive ability apparently

Table 7. *Criteria that may be useful in guiding future efforts to (a) reliably identify g/G in nonhuman animals and (b) evaluate whether a statistically identified g/G captures intelligence in a broad sense: that is, reasoning ability and behavioral flexibility. The last two columns indicate to what extent corresponding criteria have been applied in rodents and primates. See text for details (the relevant sections are indicated in italics).*

Criteria for Future Studies	Already Applied in	
	Rodents? (*g*)	Primates?
(a) To avoid statistical and methodological artifacts:		
• Use of large samples and diverse tasks, and analytical routines that do not require an a priori categorization of tasks into domains (*2.4.1*)	mostly (least for task diversity)	partly
• Replication of results in independent samples (particularly when large samples are not available, *2.4.1*)	yes	*g*: partly[1] *G*: yes
• Empirical control for confounds such as motivation, anxiety, or lower-level biological properties (*2.4.2*)	yes	no
• Avoidance of tasks prone to secondary modularization (*2.4.3*)	yes	No
(b) To explore whether an empirical finding of *g/G* captures intelligence as broadly defined (*2.5*):		
• Is *g/G* correlated with independently assessed executive functions?	yes	*g*: no *G*: yes
• Does executive function training, in contrast to a non-cognitive control training, increase *g*?	yes	no
• Is *g/G* correlated with brain size?	yes	*g*: no *G*: yes
• Is there evidence that *g/G* has been selected for?	no	yes
• Is *g/G* correlated with the ability to transfer solutions across domains (i.e., cognitive flexibility)?	no	no
• Does *g* predict performance in very difficult tasks?	yes	no
• Does *g* predict success in real life?	no	no

[1] But not successfully: see Herrmann et al. (2010b); Hopkins et al. (2014); Woodley of Menie et al. (2015)

evolved in at least some lineages, or perhaps even in birds and mammals in general, even though its evolution has had to overcome more obstacles compared to the emergence of domain-specific cognitive adaptations. Hence, the goal of this section is to delineate the conditions favoring the evolution of general intelligence.

3.1. General intelligence as response to domain-specific selection pressures

The most common approach to explain variation in cognition across species, which has a long and venerable tradition, is to look for specific cognitive challenges in the social or ecological environment and investigate to what extent species facing these challenges have evolved bigger brains (Dunbar & Shultz 2007a; 2007b; Holekamp et al. 2015; Humphrey 1976; Jolly 1966; Parker 2015; Parker & Gibson 1977). Comparative analyses, particularly in primates, have shown that brain size is indeed correlated across species with various social and ecological variables, such as social complexity based on bonded relationships (Dunbar 1992; Dunbar & Shultz 2007b) and tactical deception (Byrne & Corp 2004), or extractive foraging (Parker 2015), manipulative complexity (Heldstab et al. 2016), and perceived seasonality (van Woerden et al. 2014; 2010; 2012). Shultz and Dunbar (2006) presented similar analyses for ungulates, with similar conclusions.

Nonetheless, much variation in brain size across species remains unexplained by domain-specific pressures (Holekamp 2007; van Schaik et al. 2012). Furthermore, not all species that excel in socio-cognitive tasks, most likely because of their complex social environment, also excel in non-social tasks and evolve big brains. Socio-cognitive abilities in hyenas, for instance, are on a par with those of the larger anthropoid primates, but there is no evidence that as in primates, this would be correlated with particularly powerful cognitive abilities outside of this domain (Holekamp et al. 2015). Likewise, callitrichid monkeys outperform their sister lineage, capuchin and squirrel monkeys, in socio-cognitive abilities, but the latter have superior physical cognition (Burkart & van Schaik 2010; 2016a; 2016b). For additional examples of how specific sophisticated cognitive skills can be achieved with very small brains, see Chittka and Niven (2009) for insects or Patton and Braithwaite (2015) for fish.

The crucial question thus is: Under what conditions do specific cognitive challenges result in an increase in general intelligence (and thus brain size) rather than in domain-specific cognitive solutions that do not require large amounts of brain tissue and do not translate into benefits in other domains too?

3.2. Direct selection on general intelligence

Some have argued that general cognitive ability is not the result of a domain-specific challenge but that it is directly selected so as to help animals cope with novel or unpredictable environments and overcome unusual or complex ecological challenges. According to this *cognitive buffer hypothesis*, large brains facilitate the construction of novel behavioral patterns through domain-general cognitive processes such as innovation and learning (Lefebvre et al. 2013; Sol 2009a). In support of this hypothesis, more innovative species tend to indeed have bigger brains – in birds (Lefebvre et al. 1997) and primates (Reader & Laland 2002) – and innovation rates in the wild are correlated with *G* across primate species (Reader et al. 2011). Furthermore, innovation rates and brain size, and thus presumably *G*, predict colonization success in birds (Sol et al. 2005), mammals (Sol et al. 2008), amphibians and reptiles (Amiel et al. 2011), and in fishes (Shumway 2008; but see Drake 2007). Furthermore, large-brained birds use more successful learned strategies to avoid collision with human vehicles on roads (Husby & Husby 2014). Finally, anthropoid primates (but not lemurs, rodents, and omnivorous carnivores) cognitively buffer environmental seasonality (S. Graber et al. in prep.; van Woerden et al. 2014). Taken together, this work convincingly demonstrates that big brains are associated with greater behavioral flexibility and higher innovation rates under naturalistic settings, and that these in turn can be beneficial for a range of species when they face novel and unpredictable environments.

What remains to be answered in light of these obvious benefits, then, is why not all species evolved bigger and more-powerful brains. It is self-evident that all extant species are clearly smart enough for their current niche, but it is equally evident that a slightly better understanding that traces of a predator actually mark its presence, a better memory for which food sources already have been visited, or better object permanence to better keep track of a disappearing prey would convey a fitness benefit relative to conspecifics. We suggest that it is unlikely that focusing exclusively on potential benefits resulting from gains in brain size will further advance our understanding of the conditions under which domain-specific pressures lead to increased general intelligence. Rather, answering this question requires a focus not only on the benefits, but also on the costs of evolving a bigger brain.

3.3. Who can afford to evolve general intelligence? Cultural intelligence

Some species have larger brains than others, which, at least in primates, is associated with higher *G*. Why did these species respond to domain-specific selection pressures with an increase in general intelligence (see also sect. 3.1), or cope with environmental unpredictability by increasing their brain and intelligence, rather than opting for alternative, domain-specific adaptations (see also sect. 3.2)?

To answer these questions, it is important to keep in mind that the conditions under which large brains can evolve are to a substantial degree restricted by their costs (Isler & van Schaik 2014). Brains are energy-hungry organs that consume a large proportion of the energy available to an organism, particularly in growing immatures (Kuzawa et al. 2014). Thus, natural selection more readily favors an increase in brain size when this leads to an increase in net energy intake, a reduction in its variance, or ideally both. Furthermore, a big brain slows down the organism's development, which means that a species' ability to slow down its life history is a fundamental precondition for its opportunity to evolve larger brain size. Accordingly, the life-history filter approach (van Schaik et al. 2012) shows that slowing down life history, and thus evolving a larger brain, is only possible for species that can increase adult survival and are not subject to unavoidable extrinsic mortality, such as high predation pressure. Isler and van

Schaik (2014) have shown that such a cost perspective can explain a substantial amount of variation in brain size across primates, and that allomaternal care plays an important role in accommodating the costs associated with bigger brains (in particular, because food subsidies by allomothers help pay for the energetic costs of the growing immatures, and because of life-history consequences; see also Burkart 2017).

Natural selection thus evaluates the net fitness benefit of a bigger brain, which also takes the costs into account. The balance of benefits and costs is critically influenced by how efficiently an individual can translate brain tissue (or general cognitive potential) into survival-increasing innovations – that is, knowledge and skills. The cultural intelligence approach stresses that species that rely more systematically on social learning are more efficient in ontogenetically constructing survival-relevant skills (Herrmann et al. 2007; van Schaik & Burkart 2011; van Schaik et al. 2012; Whiten & van Schaik 2007) because social influences are very powerful domain-general canalization processes (as highlighted in Table 1). Whereas in the human literature, many approaches stress the importance of social inputs in the development of intelligence (e.g., Moll & Tomasello 2007; Tomasello 1999), the evolutionary version of this approach that suggests that social learning also plays a crucial role for the evolution of intelligence and brain size has received far less attention. Importantly, it builds on a broad notion of social learning (Heyes 1994, Box 1984; see also van Schaik et al. [2017], for a classification of social learning particularly suitable to test the predictions of the evolutionary dimension of the cultural intelligence hypothesis).

Consistent with the cultural intelligence approach, empirical results show that innovation rates in birds and primates are not only correlated with brain size or *G*, but also with the efficiency of social learning (Reader 2003; Reader et al. 2011). According to the cultural intelligence hypothesis, this is the case because, for species engaging systematically in social learning, additional brain tissue translates more reliably in survival-relevant skills, which lowers the threshold for evolution to favor an increase in brain size and general cognitive ability compared with species that do not rely on social learning. The frequency of opportunities for social learning is thus part of the answer why some lineages did evolve bigger brains, whereas others did not, even though they would all benefit from being more intelligent (van Schaik & Burkart 2011). Put in other words, we can use the pattern of solutions to the canalization problem (outlined in Table 1) to better understand under what conditions a species responds to a domain-specific selection pressure with a domain-general adaptation rather than with a narrow, domain-specific modular adaptation. The core message (to be derived from Table 1) was that all identified canalization problems can readily be overcome by social learning, and, therefore, species able to rely more on social learning should be more likely to be able to evolve domain-general cognitive adaptations. In sum, the cultural intelligence approach seems to best accommodate the findings of general intelligence as reviewed previously. For a more detailed comparison and discussion of the different approaches, see van Schaik et al. (2012) and Burkart (2017).

The cultural intelligence hypothesis was originally developed to explain why humans have evolved far bigger brains and far greater intelligence than other great apes. Tomasello (1999; see also Herrmann et al. 2007) stressed that humans have evolved a set of species-specific socio-cognitive skills that facilitate social transmission, by allowing us to participate and exchange knowledge in cultural groups from an early age on. In other words, humans have become specialized in making use of social inputs to ontogenetically construct their skills, and rather than having evolved predominantly into a "cognitive niche" (Pinker 2010), they have evolved into a "cultural niche" (Boyd et al. 2011). Our extreme dependence on the socially guided ontogenetic construction of skills can also explain why the intraspecific link between *g* and brain size within humans is relatively weak (Muthukrishna & Henrich 2016; Pietschnig et al. 2015).

Humans can thus be seen as a special case of cultural intelligence, due to the active involvement of caretakers and the improved imitative abilities of our species. This view is consistent with approaches to human cognitive evolution that stress the role of allomaternal care, which not only results in energy subsidies to growing immatures but also increases the scope of social learning through the availability of more, and more tolerant, role models, who eventually also engage in teaching (Burkart & van Schaik 2016a; 2016b; Burkart et al. 2009; Hrdy 2009; Isler & van Schaik 2014; Kline 2015).

4. Discussion

4.1. Preliminary synthesis

The current body of evidence reviewed in this article is arguably most consistent with general intelligence not being unique to humans but also present in other species, even though much validation remains to be done, as outlined in sections 2.4 and 2.5. At present, the best-supported model for both animals and humans therefore views the ecological and social cognitive skills that can be measured in a species as the result of two pathways (indicated in Fig. 3).

In the downward pathway, cognitive skills result from general intelligence, which shows strong empirical correlations with brain size and executive functions. These skills correspond to Cattell's (1963) crystallized intelligence and Geary's (2005) secondary learning. In this case, the cognitive skills are ontogenetically constructed, facilitated by mechanisms of adaptive canalization beyond Fodorian modularity (summarized in Table 1) and eventually may become automatized (secondary modularization, which makes these skills particularly difficult to identify). As stressed by the cultural intelligence approach, social learning is a particularly efficient mechanism of ontogenetic canalization, particularly in large-brained animals. In the upward pathway, cognitive skills directly emerge as a result of dedicated, Fodorian cognitive modules that are independent of general intelligence, executive functions (Firestone & Scholl 2015), or brain size.

These two pathways to cognitive skills can coexist, and in fact almost certainly do. This situation has major implications. First, closely related, big-brained species that rely to some significant extent on the downward pathway and thus general intelligence may nevertheless exhibit rather distinct social and ecological skill sets. Some of their

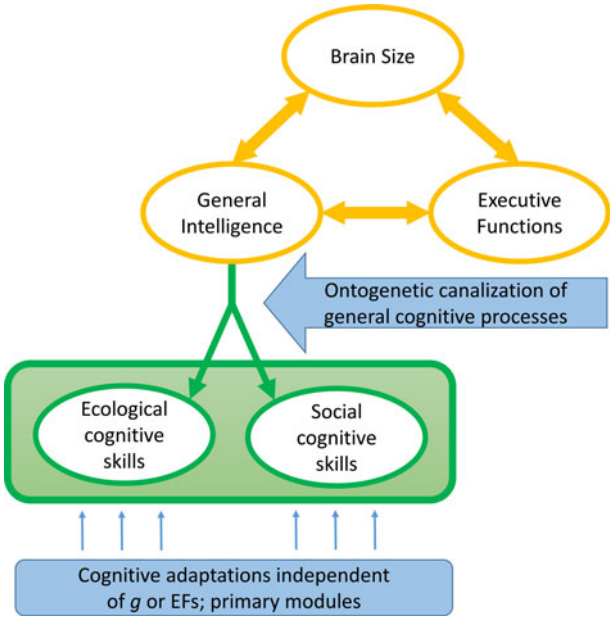


Figure 3. The origin of cognitive skills. Green: Ecological and social cognitive skills that can be measured in animals and that are visible to natural selection because they can result in fitness benefits. Yellow: Empirical interrelations between brain size, general intelligence, and executive functions. The latter two entities are only visible to selection to the extent that they are translated into fitness-enhancing cognitive skills. Blue: Adaptive canalizations that either guide the ontogenetic construction of cognitive skills from general intelligence (see Table 1 for details) or represent fully fledged Fodorian, modular adaptations that are independent of the domain-general (yellow) nexus.

species-specific ontogenetic canalization mechanisms can result in species differences in performance in specific domains, such as extractive foraging and tool use. Second, species may not primarily vary with regard to whether they have *g*, but with regard to the relative importance of these two pathways in building their skill sets, consistent with the increasing evidence for *g* in several nonhuman species. Approaches like cultural intelligence and the expensive brain framework delineate the conditions under which one or the other is more likely to evolve. This model is thus consistent with the broad pattern of results summarized in this review, including the results of mixed-species studies (sect. 2.3), and also with the idea of evolutionary continuity.

This preliminary synthesis suggests there is an alternative way of estimating the importance of general intelligence in a given species. Rather than exclusively relying on comparing the percentage of variance in performance explained by *g* (which in fact may be misleading, under the conditions outlined in sect. 2.4.1), one may attempt to estimate the importance of one pathway over the other in constructing an individual's skill set. To do so, it is crucial to be able to distinguish the origin of the skills in the green box (see Fig. 3), whether they result from the upward or the downward pathway. This is particularly difficult because eventually, skills constructed via the upper pathway may become automatized, and thus difficult to distinguish from primary modules (see also Table 2 in sect. 1.2.3). To identify them, one needs to show that they critically rely on EFs (see also Table 6b) and show signs of being effortfully learned (see also Meulman et al. 2013; Schuppli et al. 2016). This is most feasible when the learning is social, either by directly recording the socially induced patterns of attention and practice (e.g., Jaeggi et al. 2010) or by interspecific cross-fostering (see section 2.4.3) where this is feasible. This alternative way of estimating the importance of general intelligence in a given species may turn out to be a promising complement to the alternatives pursued in nonhuman intelligence research so far.

4.2. Conclusions

Overall, the body of evidence from comparative studies lends increasing support to the notion that general intelligence is not unique to humans but also present in nonhuman animals and thus is not as tied up with language as some have suggested. Intraspecific evidence for *g* is particularly strong in rodents, whereas interspecific evidence (*G*) finds most support from primate and bird studies. Nevertheless, the rather young field of research into animal general intelligence still needs to mature just as work on human intelligence has taken decades to mature.

This enterprise can obviously profit from better integrating knowledge accumulated in the longstanding tradition of human psychometrics, not only with respect to the methodological aspects highlighted previously, but also to conceptual issues. For instance, obvious parallels exist between investment theory (Cattell 1987) and cultural intelligence approaches; pursuing them further may lead to novel insights. In other domains, however, superficial similarities are misleading. Modern massive modularity, for instance, based on very broad notions of modularity and inspired by evolutionary biology (Barrett 2015), hardly informs the debate about whether general intelligence exists in nonhuman animals. Among nonhuman animals, the ancestral state most likely corresponds to animal minds being made up entirely of dedicated modular adaptations (Shettleworth 2012a; 2012b). Among extant species, the question is, which behaviors are (still) regulated this way?

It is worth emphasizing that fruitful inputs can flow in the other direction too. For instance, the availability of valid animal models of general intelligence increasingly allows studying the underlying neurobiological and genetic mechanisms in ways that would not be possible in human studies (reviewed in Galsworthy et al. 2014; Matzel et al. 2013; Plomin 2001). Furthermore, via animal studies we can experimentally address the role of factors such as exploration tendency, known to be linked to *g* in mice (Grossman et al. 2007; Light et al. 2008), most likely via mechanisms stressed by investment theory (Cattell 1987). Finally, comparative studies are indispensable in addressing the broader question of where, why, and how *g* evolved. Among the most promising evolutionary explanations for general intelligence is the cultural intelligence approach, which predicts the coevolution of social learning and general intelligence. This perspective is strongly supported by interspecific studies where social learning, but also other social abilities such as deception, are strongly correlated with *G* across species (e.g., Reader et al. 2011) and where brain size is linked to opportunities for social learning during development (van Schaik et al. 2012).

A final issue concerns both animal and human studies. In most intraspecific studies, socio-cognitive tasks were not

part of the test battery, but where they were, the results were inconclusive. Thus, whereas Hopkins et al. (2014; Woodley of Menie et al. 2015) found socio-cognitive abilities loading on *g* in chimpanzees, Herrmann et al. (2010b) did not, neither in chimpanzees nor children. This may be because the intraspecific measures of socio-cognitive abilities used so far are less suitable than interspecific ones, for instance, because they sometimes produce ceiling or floor effects. However, human test batteries typically also do not include social cognition, and the relationship between general intelligence and socio-cognitive abilities in humans therefore remains poorly understood (Korman et al. 2015). Investigating the link between socio-cognitive abilities and general intelligence within humans thus is an important research priority.

ACKNOWLEDGMENTS

We would like to thank Sonja Koski and Kay Holekamp for discussion, Sandra Heldstab for performing the PGLS analyses for Figure 2, and Jonathan Evans, Barbara Finlay, David Geary, and Willem Frankenhuis, as well as the anonymous reviewers for their valuable and very constructive feedback. This work was supported by Swiss National Funds projects 130383 and 160363 to JB, and by the A. H. Schultz-Stiftung.

Open Peer Commentary

Coexistence of general intelligence and specialized modules

doi:10.1017/S0140525X16001576, e196

Federica Amici,[a-c] Josep Call,[d,e] and Filippo Aureli[f,g]

[a]*Institute of Biology, Faculty of Bioscience, Pharmacy and Psychology, University of Leipzig, 04103 Leipzig, Germany;* [b]*Department of Primatology, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany;* [c]*Department of Psychology, University of Bern, 3012 Bern, Switzerland;* [d]*Department of Comparative and Developmental Psychology, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany;* [e]*School of Psychology and Neuroscience, University of St Andrews, St Andrews Fife KY16 9JP, United Kingdom;* [f]*Instituto de Neuroetologia, Universidad Veracruzana, 91190 Xalapa, Veracruz, Mexico;* [g]*Research Centre in Evolutionary Anthropology and Palaeoecology, Liverpool John Moores University, Liverpool L3 3AF, United Kingdom.*

amici@eva.mpg.de http://www.eva.mpg.de/pks/staff/amici/index.html
f.aureli@ljmu.ac.uk https://www.ljmu.ac.uk/about-us/staff-profiles/faculty-of-science/natural-sciences-and-psychology/filippo-aureli
call@eva.mpg.de http://www.eva.mpg.de/psycho/staff/josep-call/index.html

Abstract: Here, we specifically discuss why and to what extent we agree with Burkart et al. about the coexistence of general intelligence and modular cognitive adaptations, and why we believe that the distinction between primary and secondary modules they propose is indeed essential.

We agree with Burkart et al. that general intelligence and specialized modules likely coexist in nonhuman animals. In mammals, similar cognitive skills have evolved independently in different phylogenetic lineages (Barton & Harvey 2000; de Winter & Oxnard 2001), suggesting the existence of independently evolving modules. These specialized modules likely reflect fitness-enhancing adaptations to specific socioecological challenges (Shettleworth 2010b). However, as Burkart et al. correctly argue, nonhuman animals also solve problems flexibly across domains – something impossible for a strictly modular brain. Therefore, general intelligence and specialized modules likely coexist, at least in mammals: Although cognitive modules are the response to domain-specific socioecological challenges (Shettleworth 2010b), general intelligence may allow behavioural flexibility across domains – something especially useful in novel or unpredictable environments (Lefebvre et al. 2013; Sol 2009a).

From a neurological perspective, general intelligence and independent domain-specific cognitive skills compatibly coexist. Some properties of the human brain (e.g., amount of grey matter, neuronal speed of transmission) affect multiple brain regions, so that performance in different domains may correlate even if cognitive processes are localized in discrete regions (e.g., Jensen 1993; Lee 2007; Pennington et al. 2000). In our view, specific cognitive processes may be localized in specific brain regions also in other mammals, whereas other properties are intercorrelated across brain regions and affect all cognitive domains. Lee (2007), for instance, proposed that more synaptic connections might enhance the overall processing power of the brain, regardless of the brain regions involved. Therefore, having specific cognitive modules and more synaptic connections are two different brain characteristics that likely coexist.

In our view, Burkart et al. downplayed the importance of multifactor (as opposed to g-based) approaches in human intelligence (e.g., Kaufman 1979; Sternberg 1988; Gardner 1993). The concept of g, originally postulated by Spearman (1927), has been challenged on countless occasions and its current use in human IQ assessment is marginal at best, having been largely replaced by multifactor theories (see Kaufman 2009). An excessive reliance on *g* prevented Burkart et al. from considering multifactor approaches that may better capture interspecific cognitive diversity without necessarily invoking modularity. Several studies in nonhuman mammals have failed to find *g* and instead support a multifactor view of intelligence (e.g., Amici et al. 2012; Herrmann et al. 2007, 2010b; Herrmann & Call, 2012; Kolata et al. 2008; Schmitt et al. 2012). We suspect that the attractiveness of *g* stems from its simplicity and its use as a bastion against radical modularity. However, a multifactor view of intelligence should not be conflated with a modular view of the mind, at least not the kind of modularity defended by some evolutionary psychologists (e.g., Cosmides & Tooby, 2002). The multifactor view is general in spirit, as its factors subserve multiple cognitive problems, but each factor is specialized in particular operations (e.g., inference) or capacities (e.g., working memory). We think that a substantial portion of interspecific (and interindividual) variation in cognition can be captured by a multifactor theory without invoking modules, and as such, the multifactor approach is more germane with the notion of *g* than that of radical modularity.

We agree with Burkart et al. that different experimental and statistical approaches may lead to different results. Thus, finding *g* may, at least partly, depend on which data are included and how they are analysed. In particular, Herrmann and Call (2012) argued that task selection may inflate the relative importance of general intelligence (a point that Burkart et al. also made) by, for instance, selecting tasks that share a key feature (e.g., associative learning). Burkart et al. also argued that the allocation of tasks to specific domains (as done in confirmatory analyses and some Bayesian approaches) may be problematic, although it is possible to limit the drawbacks of a priori allocation by selecting multiple basic tasks with low cognitive demands (see Amici et al. 2012). Meta-analyses based on large data sets are especially useful for large-scale interspecific comparisons, but they often entail missing information (e.g., no interindividual variation), rely on data that are not evenly distributed across species, and disregard potentially important methodological differences across studies. These problems remain a challenge for future research, also because it is not easy to conceive tasks in which single cognitive skills are required.

We thought that the distinction between primary and secondary modules was useful. Burkart et al. argue that, through ontogeny, individuals may specialize in a certain domain, learning specific skills that become automatized and therefore appear to be domain-specific, even if they are not. The experimental distinction between primary and secondary modules is not easy, and relates to the more general problem of disentangling the relative contribution of evolutionary forces and developmental experience to cognition. Although the epigenesis of cognitive skills in nonhuman mammals is still largely unexplored, cross-fostering experiments would be a powerful tool to differentiate between evolutionarily selected and developmentally acquired behaviour. Experimental studies have shown that young macaques change their reconciliation tendencies (which are usually considered species-specific) depending on the social context in which they are raised (de Waal & Johanowicz 1993). Evolutionary forces and developmental experience are intertwined in complex ways: Differentiating between primary modules and ontogenetically acquired skills is an essential point that future research will need to address.

Finally, concerning the relative contribution of general intelligence and primary modules across taxa, there are various hypotheses as to how they should vary. On the one hand, an ecologically oriented approach suggests that taxa living in more unpredictable environments could especially benefit from behavioural flexibility across domains, and thus more strongly rely on general intelligence (Lefebvre et al. 2013; Sol 2009a). On the other hand, a more socially oriented approach suggests that taxa showing social learning can more efficiently acquire relevant skills through ontogeny without having to mainly rely on cognitive modules for their survival (Herrmann et al. 2007; van Schaik & Burkart 2011; van Schaik et al. 2012). Future research will need to find creative ways to contrast these hypotheses, while controlling for the existence of secondary modules.

An all-positive correlation matrix is not evidence of domain-general intelligence

doi:10.1017/S0140525X16001990, e197

Rosalind Arden[a] and Brendan P. Zietsch[b]

[a]*London School of Economics & Political Science, London, WC2A 2AE, United Kingdom;* [b]*University of Queensland, Brisbane, QLD 4072 Australia.*

r.arden@lse.ac.uk **zietsch@psy.uq.edu.au**

Abstract: We welcome the cross-disciplinary approach taken by Burkart et al. to probe the evolution of intelligence. We note several concerns: the uses of *g* and *G*, rank-ordering species on cognitive ability, and the meaning of general intelligence. This subject demands insights from several fields, and we look forward to cross-disciplinary collaborations.

Burkart et al. make a substantial contribution to the literature on the evolution of intelligence. We agree with the implicit view of the authors that fostering connections between contiguous fields is essential in working towards a comprehensive understanding of intelligence. The shared goal includes identifying, at least, the selection pressures that shaped cognitive abilities in different species; the structure of cognitive abilities within different species; outcomes associated with intelligence; and the genetic architecture of intelligence. The target article helpfully reaches out to engage with scholars, questions, and methods emerging from several fields including comparative and differential psychology. We value highly this drawing together of disciplines. Here we raise some points arising from Burkart et al.'s work.

First, we do not find compelling the authors' argument that positive correlations among different cognitive abilities – and the resulting latent variable *g* – reflect domain-general intelligence. ("[E]vidence for domain-general intelligence in humans, estimated by the first factor derived in psychometric, factor-analytical approaches, is pervasive …" [sect. 1.1.2, para. 3].) By contrast, our empirically testable prediction is that positive correlations among cognitive abilities, and the resulting *g* factor, will be found within most animal species, whether the species exhibit domain-general intelligence. We expect this because random alterations to a complex system usually degrade its function; genetic mutations that affect multiple cognitive abilities will tend to affect them in the same direction (i.e., deleteriously). Such directional pleiotropy would cause positive correlations among cognitive abilities (even within species that do not exhibit domain-general intelligence). This conjecture is supported by the finding that lower scores on cognitive tests are linked with a greater proportion of the genome in runs of homozygosity (a measure of the extent to which recessive alleles are expressed) (Howrigan et al. 2016).

Likewise, *G* – a latent variable arising from factorial analysis of task scores *between* species – need not reflect domain-general intelligence. In the absence of domain-general intelligence, between-species differences in brain size, neural integrity, complexity, or myelination, for example, could affect different cognitive abilities in the same direction, leading to *G*. Therefore, neither *g* nor *G* is evidence for domain-general intelligence. Further, the causes of *g* and *G* may be unrelated; *g* might be caused by directional pleiotropy, but *G* could not be. We agree that existing within-species psychometric studies are few, small, and underpowered. The cure is larger studies.

Another important point is that, because latent variables are by definition unobservable, neither *g* nor *G* can itself be a direct target of selection – contrary to Burkart et al.'s suggestion that "*G* is thus the principal locus of selection in the evolution of primate intelligence" (sect. 2.2, para. 4); *g* or *G* may reflect a real trait that is visible to selection (Borsboom & Dolan 2006), but we know of no conclusive evidence on this. Identifying biological or cognitive correlates of *g* and *G* is a useful approach to this question, but correlation is not causation, and so the cause(s) of *g* and *G* remain unclear. An additional note on the topic of selection is that, contrary to the target article (and Woodley et al. 2015), greater heritability does not indicate stronger recent selection – in fact, all else being equal, the opposite is true (Fisher 1930).

A linked issue is that the nature and cause(s) of *g* and *G*, and their relation to natural selection, depend on the tasks that are used to derive the factors. For example, interspecies differences in performance on behavioural tasks may depend on the match of the tasks to the species' typical environments and physical abilities as well as to their cognitive abilities (Barrett 2011) – in which case, the cause(s) of *G* could have environmental, physical, or cognitive sources.

Also, probing *G* does not answer the question "why are some species better at 'catching on' more generally than others?"; the answer to that lies in the recurrent problems posed by different ecologies and the costs and benefits of solving them. The costs of "generalising" make relatively more domain-general brains a better deal in some settings than in others. We should be cautious in rank-ordering intelligence between species, especially in the absence of comprehensive descriptions of cognitive abilities at the within-species level. Although it is manifestly true that some species are generalisers more than others (compare, for example, koalas with racoons), it is also the case that a smart elephant makes a lousy bat.

It should be noted that even human intelligence, which has been shaped by selection, is not completely general; it is better described as under-specified. For example, although we may inhabit a 10-dimensional universe (Green & Schwarz 1984), we are unable to form a mental image of higher dimensional figures because our minds have evolved in a space containing relevant objects of only three or fewer dimensions.

We note that we can learn much about the evolution of intelligence from genetic analyses of cognitively well-characterised populations including parameters such as heritabilities, genetic correlations (among mental traits and biological substrates within species), and coefficients of genetic variation. Genetic

studies will allow us to test relations among any observed *g* factors and other fitness-related traits, and to explore evolutionary questions concerning convergence and homologies of cognitive abilities, or mechanisms that contribute to them, across species.

Last, we urge upon us all, conscious perspective-taking of those in other fields. We are all "cursed with knowledge" (Pinker 2014, p. 11). Unpalatable phrases like "positive manifold" (e.g., sect. 1.1.1, para. 3) and "phylogenetic inflection" (sect. 1.2.2, para. 4) act as caltrops impeding the free flow of knowledge and scholarship across disciplines. Reviewers and journals can help by emphasising writing clarity. In saying this, we are not criticising the target article but celebrating and promoting the shared mission to help scholars talk to one another effectively. The focal problem, understanding the evolution of intelligence, is hard; maximising bandwidth across fields is essential.

Negative results are needed to show the specific value of a cultural explanation for *g*

doi:10.1017/S0140525X16001588, e198

Ruben C. Arslan,[a,b] Christoph J. von Borell,[a,b,e] Julia Ostner,[b,c,d] and Lars Penke[a,b]

[a]*Department of Biological Personality Psychology, Georg Elias Müller Institute of Psychology, Georg August University Göttingen, 37073 Göttingen, Germany;* [b]*Leibniz ScienceCampus Primate Cognition, 37077 Göttingen, Germany;* [c]*German Primate Center Leibniz Institute for Primate Research, 37077 Göttingen, Germany;* [d]*Department of Behavioral Ecology, Johann-Friedrich-Blumenbach Institute for Zoology and Anthropology, Georg August University Göttingen, 37073 Göttingen, Germany;* [e]*Research Training Group 2070 – Understanding Social Relationships, Georg August University Göttingen, 37073 Göttingen, Germany.*

ruben.arslan@gmail.com
https://www.psych.uni-goettingen.de/en/arslan
christoph.borell@psych.uni-goettingen.de
https://www.psych.uni-goettingen.de/en/biopers/team/borell
julia.ostner@biologie.uni-goettingen.de
https://www.uni-goettingen.de/en/164051.html
lpenke@uni-goettingen.de
https://www.psych.uni-goettingen.de/en/biopers/team/penke

Abstract: Burkart et al. suggest that social learning can explain the cognitive positive manifold for social animals, including humans. We caution that simpler explanations of positive trait intercorrelations exist, such as genetic load. To test the suggested explanation's specificity, we also need to examine non-social species and traits, such as health, that are distal to cognitive abilities.

We commend Burkart et al. for writing a clear review of the available data on intraspecies *g* and interspecies *G*. Although data on individual differences on nonhuman animals are scarce, mapping out the potentially far-ranging implications will, we hope, encourage more high-quality nonhuman individual differences research. The authors' effort to establish clearer and cross-species terms in the modularity debate and to highlight the existence of secondary modules is very welcome. Together with the acknowledgment, as nicely laid out by Burkart et al., that evolution is a tinkerer, not an engineer (Jacob 1977), we think these efforts will lead to progress in the understanding of the evolution and genetics of general intelligence.

We agree that integration between subdisciplines currently focused either on universals or individual differences (e.g., evolutionary psychology and behavior genetics) is necessary (Arslan & Penke 2015; Penke et al. 2007) and hope warnings about the lack of generalizability in nonhuman animal cognition research will be heeded (Arden et al. 2016; Thornton & Lukas 2012). On the other hand, we are unsure whether the effort to connect the appearance of *psychometric g* to the presence of *general intelligence* in a species succeeds. This leads us to examine what could cause *g* in the absence of a core intelligence trait.

We agree with the authors that psychometric *g* and domain-general intelligence should not be thoughtlessly equated (Penke et al. 2011). Although a species whose individuals vary in a core domain-general intelligence ability should exhibit a psychometric *g*, finding a psychometric *g* does not imply that variation in a core ability causes it. If *g* were caused by a core ability, then training this ability should show transfer effects to distal cognitive abilities. The authors cite such training studies, but acknowledge controversy about bias and methods (see also Colom et al. 2013; Redick 2015). Noack et al. (2014) concluded that the existing literature cannot establish such latent transfer effects. Claims of bilingual advantage have been similarly contested (Paap et al. 2015). If training of purported core abilities such as executive functions does not increase latent *g*, the case for core abilities causing the positive manifold weakens considerably. Moreover, positive correlations have not only been found among cognitive abilities but also between cognitive abilities and other fitness-related traits such as health, psychopathology, and height (Arden et al. 2016). Hagenaars et al. (2016a) showed molecular evidence that a shared genetic aetiology underlies the phenotypic associations between health and intelligence.

Reasonable, less cognition-specific explanations have been put forward to explain such positive manifolds. First among them is probably genetic load (Hill et al. 2016a; Penke et al. 2007). Individuals vary in the number of deleterious genetic mutations they carry. Depending on where they occur, such wrong turns on the genotype-phenotype map could affect the integrity and condition of the whole organism, its brain, or more specific abilities. If many of the variants affecting cognitive abilities are pleiotropic (be that because they affect early development steps or because they disrupt frequently re-used genetic patterns), then we would also expect positive correlations between cognitive abilities and other fitness-related individual differences (Deary 2012; Houle 2000). Although causal inference from genetic correlations is hard (Johnson et al. 2011; Solovieff et al. 2013), we ought to consider the possibility that biological pleiotropy, not just health behavior, explains associations between intelligence and health (Hagenaars et al. 2016b). For example, evolutionarily conserved genomic regions are strongly enriched for genetic variants affecting intelligence (Hill et al. 2016a). We think this can explain part of the *g* phenomenon. An explanation based on genetic load can even explain correlations between abilities resulting from primary modules. Other sources of individual differences such as stochastic events in early development can take an explanatory role similar to genetic load, by affecting early developmental steps and pervasive building blocks of the organism (Deary 2012).

Hence, our null hypothesis should not be complete independence of cognitive abilities, even if we knew they were primary modules. Some intercorrelation should be expected. The expected *degree* of intercorrelation depends on many unknowns, among them the degree of pleiotropy, the mutational target size, metabolic costs, and ontogeny of cognitive abilities.

With the nonhuman data available so far, we see a gap in the authors' case: the absence of clear negative results in the search for *g*/*G*. The authors report no taxon where the search for the g factor was conducted with sufficient power and appropriate methods but still failed. We suggest that only after also gathering data from less social species can we ask whether, for example, social learning increases correlations between cognitive abilities. And only when correlations with non-cognitive abilities are compared can we ask what explains the increased correlation between cognitive abilities. As the authors point out, nonhuman research can help test explanations for *g* with designs infeasible in humans, such as cross-fostering experiments. This extends to genetic load. Using genetically uniform strains and mutation accumulation lines (although these are already extremely time-consuming in micro-organisms) could help clarify the involvement of genetic load.

We lack the space to fully address alternative explanations for interspecies *G*, but hope to also see joint phylogenetic tree analyses of sociality and variance explained by *g*. To be able to test this, recommendations for increased sample sizes in such studies should be followed (Thornton & Lukas 2012). Differential measurement error across subtests and species has to be modelled and corrected for, not just used to explain negative findings.

In conclusion, we would add the following to the authors' call for research: We need individual differences of data along the whole gradient of sociality including maybe even octopuses, and studies should also examine more distal traits such as health and size. Then, the specific added value of the proposed model can be tested. We want to echo this and previous calls (Arden et al. 2016; Thornton & Lukas 2012) for more individual differences research on nonhuman animal cognition. We hope for more stimulating evolutionary theorizing on individual differences, as in this target article.

G but not *g*: In search of the evolutionary continuity of intelligence

doi:10.1017/S0140525X1600159X, e199

Moran Bar-Hen-Schweiger,[a] Avraham Schweiger,[b] and Avishai Henik[c]

[a]*Department of Psychology, Ben-Gurion University of the Negev, Beer-Sheva, Israel 84105;* [b]*Department of Behavioral Sciences, Academic College of Tel Aviv, Israel 61083;* [c]*New York City Medical & Neurological Offices, New York, NY 11373*

henik@bgu.ac.il http://www.bgu.ac.il/~henik
moranbar@post.bgu.ac.il schweige@mta.ac.il

Corresponding author: Avishai Henik

Abstract: Conceptualizing intelligence in its biological context, as the expression of manifold adaptations, compels a rethinking of measuring this characteristic in humans, relying also on animal studies of analogous skills. *Mental manipulation*, as an extension of object manipulation, provides a continuous, biologically based concept for studying *G* as it pertains to individual differences in humans and other species.

Burkart et al.'s review of studies involving *g* and *G* in animals is illuminating. However, the authors seem to assume that a century of studies settled the question of *g* in humans. In this commentary, we challenge this assumption. We suggest that the definition of intelligence (Gottfredson 1997) cited by the authors seems to be overly anthropocentric: It emphasizes skills characteristic of Homo sapiens. This very definition appears to constrain *g* in humans and, in our opinion, limits its generalizability in a biological context. Furthermore, the old debate about the cultural fairness of intelligence testing is testimony to the vicissitudes of defining and measuring *g* in humans.

As an alternative, we propose a description of intelligence in a biological context, inspired by the work of Piaget, who suggested that "Intelligence is an adaptation … The organism adapts itself by materially constructing new forms to fit them into those of the universe, whereas intelligence extends this creation by constructing mental structures which can be applied to those of the environment" (Piaget 1952, pp. 3–4). Thus, intelligence is the ability of a species to adapt flexibly to many environmental challenges in the service of survival. Note that this definition is species relative: It follows naturally that the larger the range of environmental conditions to which an organism can potentially adapt, the more intelligent it might be relative to other organisms with more limited repertoire of adaptations (Piaget 1971).

Furthermore, we propose that intelligence is not a trait. Rather, it is the inference by the human observer in the face of increasing the potential and scope of domain-specific skills developed by species in adapting to a variety of environmental pressures. These domain-specific skills allow for increasing the range of environments to which the organism is able to respond efficiently. Of course, the converse is also the case: Environmental changes will result in adaptation by the emergence of new domain-specific skills. Such increase is the consequence of evolving ever-larger brains, especially frontal lobes, which enables the development of an ever-greater repertoire of skills (Parker & Gibson 1977). Moreover, a procedure or skill developed specifically within a specific domain might become accessible to systems or brain structures that serve other domains (Anderson 2010). Such change in accessibility creates domain-general skills or procedures, thus improving intelligence (or adaptability) (Rozin 1976). The result appears to be *g* to the human observer. Lest we revert again to an overly anthropocentric view, we emphasize that increasing brain size is just one means of increasing survivability.

Instead of the mysterious *g* factor, we propose that *G* is reflected in a capacity: object manipulation in various species, with its evolution into a mental manipulation (MM) in humans – the hallmark of human activity. MM can be investigated by various tests using verbal, mathematical, or spatial manipulation of contents. These tests tend to correlate positively not because they reflect *g*, but because they reflect MM. We suggest that MM is the ability to perform transformations on concrete and abstract objects (e.g., mental rotation) and imagine the results, without needing the actual objects. This ability clearly improves adaptability to a wide range of environments. One example of MM is when a child learns to consider a situation from the perspective of another person. We claim that linguistic construction, as well as other cognitive processes, involves MM, so that it may be considered as an overarching principle of human operations and as the basis of human culture.

To illustrate the biological continuity and the development of domain-specific skill, we consider the ontogeny of mathematical skills in humans (described originally by Piaget 1971). Initially, babies develop the concept of *one*, *few*, and *many*, requiring direct perception of objects. Animals exhibit number concept at this level (Pahl et al. 2013). Later, children learn that abstract symbols represent quantities, and they learn how to manipulate them. Next, algebra supplants numbers at ever higher levels of abstraction, with ever more abstract manipulations (i.e., operations). An analogous analysis was offered by Greenfield (1991) regarding the development of linguistic structures from motor schemata in children. The essence of these developmental achievements is that they reflect the ability to perform transformations, translations, recombinations, projection, predictions, and so on, in infinite ways. What is crucial for adaptation is the ability to entertain the results of these MMs, and then select only the best one for action.

One challenge to which MM could offer a positive contribution is in measuring the nebulous *g*. Many, if not all, IQ testing instruments may be viewed as assessing domain-specific abilities. To what extent do they reveal an underlying, domain-general or universal ability? For example, the Raven Progressive Matrices is commonly used as a measure of the *g* factor (Deary et al. 2010). However, in a wider cultural context, this test may measure no more than a domain-specific, culturally acquired skill (Owen 1992). We suggest that MM is such an overarching set of operations. It is possible that MM started as an ability to manipulate or view actual objects designed for the visual-spatial domain. These visual-spatial specific abilities evolved to serve other domains (e.g., language) and have become accessible to other systems that serve other commitments.

In sum, we propose that the alternative conception of intelligence as offered here, compels rethinking *g* in humans. It is suggested that animal behavior, specifically, object manipulations, and perspective taking (a variant of MM) with increasing cortex, provide specific precursors to human abilities, as reviewed by Burkart et al. A good example of a transitional stage to MM demonstrating biological continuity is deception in apes (Byrne & Corp 2004).We further submit that MM may better serve as a biologically based concept for studying individual differences in humans, while providing for continuity across species.

Domains of generality

doi:10.1017/S0140525X16001606, e200

Andrew Buskell[a] and Marta Halina[b]

[a]*London School of Economics, Department of Philosophy, Logic and Scientific Method, London School of Economics, London, United Kingdom, WC2A 2AE;* [b]*University of Cambridge, Department of History and Philosophy of Science, University of Cambridge, Cambridge, CB2 3RH, United Kingdom.*

a.buskell@lse.ac.uk mh801@cam.ac.uk

http://people.ds.cam.ac.uk/mh801/

Abstract: We argue that general intelligence, as presented in the target article, generates multiple distinct and non-equivalent characterisations. Clarifying this central concept is necessary for assessing Burkart et al.'s proposal that the cultural intelligence hypothesis is the best explanation for the evolution of general intelligence. We assess this claim by considering two characterisations of general intelligence presented in the article.

Recent studies suggest that general intelligence is not limited to humans, but can be identified in a number of nonhuman species. Such studies provoke the question: How does general intelligence evolve? Burkart et al. argue that the cultural intelligence hypothesis is the best explanation for the evolution of human and animal general intelligence. According to this hypothesis, the access to socially maintained knowledge generates selection pressures for increased reliance upon domain-general capacities. In order to assess the cultural intelligence hypothesis, one must have a good grasp of what is meant by general intelligence. The authors are quick to note that this is a tricky concept to pin down. Defining general intelligence in terms of specific measures of intelligence like problem solving and learning is problematic insofar as these skills can result from adaptive domain-specific modules. The authors thus characterise general intelligence as a domain-general ability best understood in contrast to the properties of domain-specific modules (sect. 1.2.1, para. 5, 6). There are many ways to make this contrast, however. As a result, the authors characterise domain-generality in a variety of ways: as phenotypic plasticity, as being non-modular in structure, as requiring learning or other processes of "ontogenetic construction" or canalisation, as involving reasoning, or as involving belief or belief-like states. The fact that domain-generality can be understood in many distinct and non-equivalent ways is worrisome insofar as different conceptualisations of domain-generality are likely to require distinct evolutionary narratives.

To take one example, domain-generality can be purchased quite cheaply if it is understood as mere phenotypic plasticity. Peter Godfrey-Smith (1996), for instance, makes a convincing case that phenotypic plasticity is selected for in heterogeneous environments – a scenario one expects to be common. To the extent that this is the case, phenotypic plasticity will be a widespread adaptive solution, seen quite deep in evolutionary history (Godfrey-Smith's central example of *Bryozoan*, sometimes known as "sea-moss," behaviour makes this point clear). One does not need to invoke the cultural intelligence hypothesis in order to purchase such flexibility.

Elsewhere, the authors characterise domain-generality as involving reasoning and belief or belief-like states. The idea here is that domain-generality can be conceptualised as the ability to use a variety of distal cues to generate mental representations, which in turn can be used to produce adaptive behaviour (sect. 1.1.2, para. 4). Of course, mental representations can be understood more or less restrictively. In some extenuated way, simple neuronal systems like those of *Caenorhabditis elegans* "represent" or register their local environment. However, it seems clear that the authors are interested in representations in a richer sense, in line with what Kim Sterelny (2003) called "de-coupled representations." These are representational states with the function of tracking features of the environment, but which are not tightly coupled to specific types of response. Such representations identify what Sterelny called "action targets" which can be acted on in a variety of different ways to satisfy goals.

De-coupled representations are an interesting evolutionary phenomenon, and one that the cultural intelligence hypothesis may get some explanatory purchase upon. De-coupled representations are the kind of psychological structure one would expect of creatures who need to rationalise and predict the thoughts of conspecifics, as well as weigh the complex tradeoffs involved in acquiring knowledge from multiple sources. However, even here we urge caution. The coleoid cephalopods (cuttlefish, squid, and octopuses) seem to display de-coupled intelligent behaviour, particularly those of the order *Octopoda*. Octopuses display sophisticated cognitive capacities including problem solving, individual recognition, and perhaps imitation (Godfrey-Smith 2013; Mather & Kuba 2013; Roth 2013). Nonetheless, octopuses are not social, often interacting with conspecifics only during mating (Roth 2013). The existence of cephalopod intelligence may thus pose a counterexample to the cultural intelligence hypothesis even when general intelligence is understood in the restricted sense of involving de-coupled representations.

The authors might respond by arguing that the evolutionary phenomenon they are attempting to describe is not merely the existence and amplification of one of the aforementioned features of domain-general cognition, but how a conglomerate of such properties came about and increased in sophistication. This conglomerate might include flexible reasoning and learning generated by de-coupled representations, in turn underpinned by the contents of an increasingly large brain. If this conglomerate really is what the authors mean by domain-generality, however, then they need to do more in order to motivate it. Recent work on grackles and New Caledonian crows, for example, shows that behavioural flexibility occurs independently of innovativeness, problem-solving ability, problem-solving speed, and brain size (Logan et al. 2014; Logan 2016a; 2016b). Given that purported features of domain-general intelligence do not always co-occur, further justification is required to ground claims that "general intelligence" is a unitary explanandum.

We suggest that the consequences of this analysis are twofold. First, we urge the authors to be clearer about the terms they use, and to operationalise them when possible. Second, the arguments presented here suggest that the cultural intelligence hypothesis may be insufficient for explaining the evolution of general intelligence, understood as involving the evolution of de-coupled representational states. Although this hypothesis may capture some directional effects in some clades, more needs to be done in order to show that it is the best explanation for the evolution of general intelligence in all clades.

Theory of mind: A foundational component of human general intelligence

doi:10.1017/S0140525X16001618, e201

David Estes and Karen Bartsch

Department of Psychology, University of Wyoming, Laramie, WY 82071

estes@uwyo.edu bartsch@uwyo.edu

Abstract: To understand the evolution of general intelligence, Burkart et al. endorse a "cultural intelligence approach," which emphasizes the critical importance of social interaction. We argue that *theory of mind* provides an essential foundation and shared perspective for the efficient ontogenetic transmission of crucial knowledge and skills during human development and, together with language, can account for superior human general intelligence.

We commend Burkart et al. for their impressive review of current theory and research on the evolution of general intelligence.

Although their analysis yields a persuasive account of the empirical data regarding general intelligence across species, it also motivates increased attention to key components of human cognition that, in our view, will prove to be essential for a complete understanding of the evolution of cognition.

A comprehensive theory of general intelligence and its evolution must somehow account not only for the commonalities evident across species, as targeted by Burkart et al., but also for the obvious wide divide between humans and nonhuman primates. Although we agree that the "cultural intelligence approach" endorsed by the authors will be at the heart of a successful theory, we believe their analysis also suggests directions for further elucidation that would specifically address what is special about human cognition.

Specifically, and consistent with the cultural intelligence approach, we propose that language and theory of mind (ToM) together can go far toward explaining superior human general intelligence. Language and ToM arguably have the markings of primary modules, at least as characterized by Burkart et al. (sect. 1.2.3, Table 2): Both appear rapidly and dependably early in development across all cultures without explicit instruction and even in the most disadvantaged and deficient environments (e.g., Wellman et al. 2001).

Others can better address the role of language in human intelligence (see Dunbar 1998; Smith 1996), so we focus here on the contribution of ToM, defined as the universal propensity of humans to understand and explain their own and others' behavior in terms of internal mental states and processes such as beliefs, desires, goals, and intentions (e.g., Wellman 1990). We note here that language and ToM are so ontogenetically and symbiotically intertwined that they are difficult to discuss separately. Indeed, some have made a strong case for their co-evolution: "The evidence at hand suggests that language and theory of mind evolved … in constant interaction, serving one primary adaptive goal: to improve social coordination" (Malle 2002, p. 280).

Nevertheless, a conceptual case can be made for ToM alone being especially foundational for human general intelligence. As a developing domain-specific causal framework that supports advanced meta- and social cognition, such as recognition of the possibility of false beliefs and variation in knowledge states between individuals, ToM contributes essentially to social learning as characterized by the cultural intelligence approach. We offer here a few examples of phenomena illustrating the crucial role of ToM in social learning and, by extension, the evolution of human general intelligence. For a recent collection of relevant research and commentary on these issues, see Legare and Harris (2016), and especially Tomasello (2016).

Even children under age 2 engage in observational imitative learning that depends on attention to the actor's intended action or goal. For instance, Meltzoff (1995) demonstrated that when shown an actor who tried and failed to achieve a goal, such as hanging a string of beads on a peg, 18-month-old children imitated what the actor *tried* to do rather than what the actor actually did (dropping the string of beads). Carpenter et al. (1998) found that young children shown adult demonstrations of action sequences containing both accidental and intentional actions more often reproduced the intentional actions, suggesting that even early observational learning is contingent on an appreciation of another's intentional state.

Similarly, research on children's early attunement to potential knowledge in others reveals the advantages for a social learner afforded by an emerging ToM. Even 2-year-olds attempting to locate a desired object efficiently seek help from informed rather than ignorant adults (O'Neill 1996), demonstrating an incipient appreciation of mental states that will eventually help them when they seek to obtain information rather than objects. Recent studies of children's reaction to testimony suggest that, although children generally assume what they are told is truthful (e.g., Harris 2012), they are more willing to believe an expert or experienced source. Moreover, as children advance from early to middle childhood, their evaluation of others' testimony increasingly recognizes the possibility of deception or distortion (Mills 2013). On the positive side, children's growing sophistication eventually includes recognizing the specifically pedagogical motives of adult teachers, as described in Gergely and Csibra's (2006; Csibra & Gergely 2009) proposed *theory of natural pedagogy*. Children's particular "readiness" to learn through social instruction (and also to teach others; see Flynn & Whiten 2008) hinges on a recognition of the internal content of others' minds. From our own research, we can add that young children's reflections on their own learning, as assessed in transcripts of at-home conversations, are in fact particularly focused on their teachers and what their teachers know as opposed to other aspects of learning events, such as how or when learning occurred (Bartsch et al. 2003).

Early childhood is replete with phenomena suggestive of the central role of ToM in social learning. In addition to these examples, the acquisition of belief understanding, widely recognized as a hallmark of developing ToM and first evident around 4 years of age in children's explicit predictions and explanations of action (e.g., Wellman et al. 2001), can also be viewed as foundational to the most advanced human achievements, such as those that occur in the collaborative activity of scientific research and the uniquely human creation of formal methods to facilitate learning (e.g., Meltzoff et al. 2009). At a fundamental level, our greatest accomplishments must rest on our basic capacity to imagine and recognize the variations and vulnerabilities of human cognitive states. With this in mind, future researchers guided by Burkart et al.'s final recommendation to explore further the relationship between intelligence and socio-cognitive abilities would do well to direct those efforts toward theory of mind.

Understanding the relationship between general intelligence and socio-cognitive abilities in humans

doi:10.1017/S0140525X1600162X, e202

Pablo Fernández-Berrocal,[a] Rosario Cabello,[b] and María José Gutiérrez-Cobo[a]

[a]*Department of Basic Psychology, Faculty of Psychology, University of Málaga, 29071, Spain;* [b]*Department of Developmental and Educational Psychology, University of Granada, Granada, 18071, Spain.*

berrocal@uma.es

https://www.researchgate.net/profile/Pablo_Fernandez-Berrocal

rcabello@ugr.es

https://www.researchgate.net/profile/Rosario_Cabello

mjgc@uma.es

https://www.researchgate.net/profile/Maria_Jose_Gutierrez-Cobo

Abstract: Burkart et al. consider that the relationship between general intelligence and socio-cognitive abilities is poorly understood in animals and humans. We examine this conclusion in the perspective of an already substantial evidence base on the relationship among general intelligence, theory of mind, and emotional intelligence. We propose a link between general intelligence and socio-cognitive abilities within humans.

Burkart et al. assess what studies on general intelligence in nonhuman animals mean for current theories about the evolution of general intelligence. Although we agree with their conclusions in favor of the cultural intelligence approach, we do not entirely agree with their assessment that the relationship between general intelligence and socio-cognitive abilities is poorly understood in animals and humans (sect. 4.2, para. 4).

In this commentary, we aim to place their conclusion in the perspective of an already substantial evidence base demonstrating a relationship between general intelligence and socio-cognitive abilities in humans. We review recent meta-analyses on this relationship, focusing on connections among general intelligence, theory

of mind (ToM; Baron-Cohen et al. 1997; Baron-Cohen et al. 2001) and ability-based emotional intelligence (EI; Mayer & Salovey 1997).

ToM is the ability to attribute mental states (e.g., emotions, intentions, or beliefs) that differ from our own (Baron-Cohen et al. 1985; Baron-Cohen et al. 2001). ToM is widely assessed using the Reading the Mind in the Eyes Test (RMET; Baron-Cohen et al. 1997; 2001), which can reveal intersubject differences in social cognition and emotion recognition across different groups and cultures (Fernández-Abascal et al. 2013). A recent meta-analysis involving 3,583 participants revealed a small positive correlation between general intelligence and RMET score (r = .24), with both verbal and performance IQ showing similar correlations with RMET score (Baker et al. 2014). The authors of that meta-analysis concluded that intelligence contributes significantly to ToM, with verbal and performance IQ contributing equally.

EI is a construct central to conceptualizing and evaluating socio-cognitive abilities. EI refers to the ability to reason validly with emotions and with emotion-related information and to use emotions to enhance thought (Mayer & Salovey 1997; Mayer et al. 2008). The most common measure of ability-based EI is the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT; Mayer et al. 2002). This test assesses the four primary abilities (branches) of the Mayer and Salovey model of EI: perceiving emotions in oneself and others, using emotions to facilitate thought, understanding emotional information, and regulating emotions in oneself and others (Mayer & Salovey 1997). MSCEIT assesses these emotional abilities by asking the subject to solve a series of emotion-based problems, thereby avoiding the high risk of bias associated with self-report EI measures. MSCEIT-based studies have demonstrated a relationship between general intelligence and EI. For instance, Webb et al. (2013) found significant correlations of MSCEIT score with general IQ, verbal IQ, and performance IQ. A meta-analysis of 53 studies involving 3,846 participants found positive correlations of scores on the MSCEIT or its forerunner MEIS with general intelligence (r = .30), verbal intelligence (r = .26), and nonverbal intelligence (r=.23) (Kong 2014).

Factor-analytic exploration of how mental abilities correlate with one another suggests an even broader range of intelligences linked to ability-based EI, including fluid intelligence, crystallized intelligence, and quantitative reasoning (Legree et al. 2014; MacCann et al. 2014). These intelligences lie within the second stratum of the Cattell-Horn-Carroll model (McGrew 2009). Further evidence for the relationship of ability-based EI with a range of broad intelligences comes from a study involving more than 12,000 people ranging in age from 17 to 76 years (Cabello et al. 2016). In this study, MSCEIT scores varied with age according to an inverted-U curve: Younger and older adults scored lower than middle-aged adults, just as reported for several other intelligences.

In this way, the extensive literature on ability EI provides substantial evidence linking various types of intelligence to socio-cognitive abilities. Nevertheless, one thing that remains unclear is how the EI assessed on the MSCEIT relates to executive functions, some of which – such as inhibitory control and working memory – strongly correlate with general intelligence, as Burkart et al. point out (sect. 1.1.2, para. 2). Gutiérrez-Cobo et al. (2016) systematically reviewed 26 studies on the relationship between EI and cognitive processes reflected in tasks such as the Stroop task or Iowa gambling task. The authors found that performance-based ability EI (such as measured on the MSCEIT) – but not self-report EI – positively correlated with efficiency on emotionally laden tasks. In contrast, no correlations were observed between EI measured in various ways and non-emotionally laden tasks. These findings suggest that the greater intelligence reflected in higher ability-based EI can mean superior performance on emotionally laden socio-cognitive tasks, but not necessarily on other kinds of tasks.

In summary, the body of studies examining ToM and ability-based EI build a strong case that general intelligence, particularly intelligence in the second stratum of the Cattell-Horn-Carroll model, is associated with socio-cognitive abilities in humans. Studies of ability-based EI and cognitive processes nuance that this relationship is likely to be complex: For example, higher ability EI may lead to more efficient cognitive processes in emotionally laden tasks but not other tasks. A link between general intelligence and socio-cognitive abilities coincides nicely with studies from affective and social neuroscience showing that emotion processing and cognition in the brain are highly intertwined and mutually determined (Phelps et al. 2014).

Future work should (1) examine to what extent different socio-cognitive abilities are related (e.g., how are ToM and EI related?), (2) analyze to what extent different socio-cognitive abilities relate to general intelligence, (3) test whether and how specific social inputs play a role during ontogenetic construction of socio-cognitive abilities, and (4) identify brain regions involved in different socio-cognitive abilities and examine their relationship and overlap with regions implicated in general intelligence.

Taking a multiple intelligences (MI) perspective

doi:10.1017/S0140525X16001631, e203

Howard Gardner

Harvard Graduate School of Education, Harvard University, Longfellow Hall 235, Cambridge, MA 02138.

hgasst@gse.harvard.edu http://multipleintelligencesoasis.org/

Abstract: The theory of multiple intelligences (MI) seeks to describe and encompass the range of human cognitive capacities. In challenging the concept of general intelligence, we can apply an MI perspective that may provide a more useful approach to cognitive differences within and across species.

In line with the view of most psychologists and psychometricians, Burkart et al. assume that there is a single general intelligence (g); controversy centers around the identity and nature of domain-specific computational capacities and the extent to which nonhuman animals can be said to have a g-like capacity.

Over the past several decades, researchers have challenged this consensus and developed alternative ways of conceptualizing human intellect (Guilford 1967; Sternberg 1984). In my case, I deliberately disregarded paper-and-pencil instruments of the sort favored in scholastic settings, and which almost always yield a "positive manifold." Instead, culling evidence from a range of disciplines – from anthropology and education to neuropsychology and evolutionary biology – I put forth the claim that human beings are better described as having a set of relatively independent computational capacities, which I termed the "multiple intelligences" (Gardner 1983/2011).

According to my analysis, the kind of intelligence typically measured in IQ tests is scholastic intelligence – the bundle of skills needed to succeed in modern secular schools. In my terms, success on such instruments depends on a combination of linguistic and logical-mathematical intelligences, with spatial intelligence sometimes tapped as well. It is worth noting that, at the extremes, strength (or weakness) with one of these intelligences does not predict comparative strength (or weakness) with the other (Determan 1993). Largely ignored in standard measures of intellect are several other intelligences that I identified: interpersonal and intrapersonal intelligences (often described as social or emotional intelligence), musical intelligence, bodily kinesthetic intelligence, and naturalist intelligence. Because we do not have comparable instruments to assess nonscholastic intelligences (but see Gardner et al. 1998), we do not know to what extent ability in, say, musical intelligence correlates with strength (or deficits) in, say, social or intrapersonal intelligence. Yet ample evidence confirms that these intelligences can be dissociated from one another, as happens with prodigies (Winner 1997) or

individuals on the autistic spectrum (Silberman 2016). Put differently, ability to succeed in school settings is decreasingly important, as one ventures to contexts that differ significantly from the canonical Western school.

Even as MI theory differs from a *g*-centric view of the world, it also differs from Fodorian modules. Intelligences may contain specific modules (e.g., linguistic intelligence may contain a parsing or phoneme discrimination module), but their exercise in the world is far less reflexive, far more adaptive. An individual skilled in linguistic intelligence is able to speak, write, communicate, and learn new languages and the like. Skill in spatial intelligence involves making sense of local two-dimensional arrays, as well as navigating around a neighborhood or, indeed, around the globe.

It may seem that the intelligences are a grab-bag of primary Fodorian modules as well as more-general secondary information-processing or problem-solving capacities processing certain kinds of content. And, indeed, as we attempt to make sense of human cognition, that characterization proves serviceable. MI theory stands out less in terms of the precision of its claims with respect to the execution of tasks in the world than in its challenge to the notion that there exists any sensible and defensible notion of general intelligence – even within Homo sapiens.

For those sympathetic to an MI view, formidable questions remain. What are the basic building blocks of intellect? To what extent is each heritable? How do strictly modular capacities interact with ones that are more permeable? Do we need to posit a separate "executive function," a so-called "central intelligence agency," or does such a capacity emerge naturally out of intrapersonal intelligence (the ability to know oneself accurately) and logical intelligence (the ability to reason about one's actions)?

An MI perspective yields far more specific pictures of how human beings carry out the raft of tasks for which the species has specifically evolved as well as those tasks that have emerged over the centuries by virtue of newly emerging cultural artifacts and technologies, and, perhaps, acts of nature (e.g., diseases, volcanic eruptions). Furthermore, such a perspective suggests an alternative approach to the issue addressed in the target article.

Instead of invoking *g*, plus specific modules, one can instead break down any task in terms of its demands on specific intelligences (e.g., playing chess involves logical and spatial intelligence but little bodily or musical intelligence) as well as the various ways in which one can become proficient at the task (e.g., some chess players weigh interpersonal intelligence – knowing the opponent – much more than do others). We avoid the conundrum that human intelligence is most naturally assessed through language-based instruments, and yet such instruments cannot be employed with other animals – leaving us with a situation where we can do species comparisons only by eliminating what is widely regarded as the essence of human intellect. By the same argument, we cannot use "musical intelligence" of birds, or the "echoing intelligence" of bats, again ignoring a dominant intellectual capacity. More generally, we may be better able to trace the similarities and differences between human beings and particular species (be they birds, bats, or dolphins) if we think of them in terms of each species' own dominant and less salient intelligences, rather than their having more or less of *g*.

Of mice and men, nature and nurture, and a few red herrings

doi:10.1017/S0140525X16001643, e204

Marc D. Hauser

Risk-Eraser, West Falmouth, MA 02574.

marc@risk-eraser.com **www.risk-eraser.com**

Abstract: Burkart et al.'s proposal is based on three false premises: (1) theories of the mind are either domain-specific/modular (DSM) or domain-general (DG); (2) DSM systems are considered inflexible, built by nature; and (3) animal minds are deemed as purely DSM. Clearing up these conceptual confusions is a necessary first step in understanding how general intelligence evolved.

"The best-laid schemes o' mice an' men," penned Robert Burns in his ode to mice. It is an astute observation of how our intelligence has outwitted theirs. Though I appreciate Burkart et al.'s attempt to synthesize a wild and wooly comparative literature on general intelligence in mice, men, and many other species, they have introduced a few red herrings and false premises that muddy the waters and undermine suggestions for future research.

The problem starts with the authors' initial premise: Scholars tend to view the mind as *either* domain-specific/modular (DSM) *or* domain-general (DG), and those who lean to DSM see the mind as predetermined and inflexible, and thus largely the work of nature. These views conflict with a theory of general intelligence. Burkart et al. claim that their framework shows "that human cognition involves elements of domain-specific and domain-general processes" (sect. 1.2.3, para. 4), and in contrast to prior views, "animal minds need not be bundles of specialized cognitive adaptations" (sect. 1.2.3, para. 4). But their premise is false as is their characterization of animal research. This commentary addresses these misconceptions and introduces some additional distinctions in order to productively explore how general intelligence evolved.

Those who have synthesized DSM perspectives (e.g., Pinker 1997) do not deny the existence or significance of DG capacities: evolutionarily ancient mechanisms that typically interface with and often constrain the outputs of each domain. Research on theory of mind (ToM), number, and language – domains often considered as modules – has long explored how executive functions interact with the computations and representations of each domain (Bradford et al. 2015; Soltész et al. 2011). For example, delays in the expression of ToM and number competence are intimately related to the development of working memory, whereas performance on ToM tasks can be improved by lifting constraints that arise from inhibitory control or perseverative responses. Thus, although it is inaccurate to pigeonhole scholars as either DSM or DG, it is true that those who have explored the nature of DSM systems are more interested in them and in how they can be characterized on the basis of evolutionary theory. Similarly, although the generative computations that subserve language competence (but also other domains such as music, number, and ToM) have no limit, our capacity to produce or comprehend sentences is limited by working memory. Thus, although DSM-focused researchers tend to emphasize the nature of the representations and computations within a domain or module, they don't deny the existence or potentially constraining impact of DG processes.

Of relevance to the evolution of general intelligence is the underlying architecture of DSM systems. Here, too, Burkart et al. mischaracterize these as innate and inflexible. Research on faces reveals this error. Neurobiological studies in macaques and humans reveals dedicated circuitry that is consistent with a DSM perspective. However, this system matures slowly over time and depends on experience with faces as elegantly demonstrated by studies of individuals with early-appearing cataracts that were later removed (Rhodes et al. 2017). A similar characterization applies to language, wherein there are core underlying computations and representations, some specific to language and others shared (Hauser & Watumull 2016), but with experience selecting among the options to generate specific languages (e.g., French, English).

Lastly, it is simply not the case that nonhuman animals are perceived as mere bundles of modules, fixed and inflexible. Research on model systems such as aplysia and songbirds reveals both ancient, general mechanisms for learning and memory, as well as highly dedicated systems that nonetheless show plasticity. For example, although passerines acquire their song on the basis of

specialized circuitry that enables vocal imitation, this same system requires specific input (e.g., species-specific song), is not engaged for other vocalizations (e.g., alarm calls), and in some species, shows plasticity throughout life as individuals create new songs each season. In addition, many researchers have recognized and detailed other DG processes that go beyond what Burkart et al. discuss. For example, there is considerable comparative work exploring the concept of "sameness," analogical reasoning, and algebraic computations (Martinho & Kacelnik 2016; Smirnova et al. 2015; ten Cate 2016). These are not part of the executive system, have not typically been linked to general intelligence, and yet they cut across domains and appear evolutionarily ancient.

Putting these strands together suggests that any approach to exploring the evolution of intelligence must consider the interaction between DSM and DG, understand the specificity of the content of DSM, examine a diversity of DG systems (i.e., beyond executive functions), and document how maturational changes in DG can impact the ontogeny of DSM. The content of a domain is particularly relevant as tasks within the general intelligence battery are often assumed to be part of a given domain without rigorous testing. Take, for example, work on tool use. Many researchers have considered tool technology a domain, one based in part on the functional design features of its objects. Thus, when animals such as chimpanzees and New Caledonia crows – natural tool users – show sensitivity to an object's design features, using those objects that are most likely to lead to successful outcomes, we consider this to be evidence of domain-specificity. And yet, cotton-top tamarins – a species that never uses tools in the wild and shows virtually no interest in object manipulation in captivity – show the same kind of sensitivity to an object's design features as chimpanzees and crows; furthermore, this sensitivity appears early in ontogeny in the absence of experience (Hauser et al. 2002a). This suggests that we should be more cautious with our claims of DSM capacities, and thus, how we classify the tasks within a general intelligence battery.

In conclusion, although Burkart et al. introduce a tension between DSM and DG that doesn't exist, incorrectly consider DSM perspectives as innate and inflexible, and falsely accuse other scholars of classifying nonhuman animals as rigidly DSM, they are correct in emphasizing the importance of looking more deeply at general intelligence in animals. Progress will depend on a clear articulation of the different skills tapped in the general intelligence battery, and standard methods that can be implemented across a diversity of species.

The evolution of general intelligence in *all* animals and machines

doi:10.1017/S0140525X16001655, e205

Kay E. Holekamp[a,b,d] and Risto Miikkulainen[c,d]

[a]*Department of Integrative Biology, Michigan State University, East Lansing, MI 48824-1115;* [b]*Ecology, Evolutionary Biology and Behavior, Michigan State University, East Lansing, MI 48824;* [c]*Departments of Computer Science and Neuroscience, University of Texas, Austin, TX 78712;* [d]*BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, MI 48824*

holekamp@msu.edu http://www.holekamplab.org/
risto@cs.utexas.edu https://www.cs.utexas.edu/~risto/

Abstract: We strongly agree that general intelligence occurs in many animals but find the cultural intelligence hypothesis of limited usefulness. Any viable hypothesis explaining the evolution of general intelligence should be able to account for it in all species where it is known to occur, and should also predict the conditions under which we can develop machines with general intelligence as well.

In their rich and thought-provoking review, Burkart et al. use impeccable scholarship to produce a heroic synthesis of multiple complex literatures. Their two main goals are to critically evaluate the question of whether general intelligence exists in nonhuman animals, and to evaluate the implications of general intelligence for current theories about the evolution of cognition. In our view, they accomplish the first goal extremely effectively, making a compelling argument that general intelligence is indeed widespread among animals. Regarding their second goal, they argue that existing data from vertebrates support the cultural intelligence hypothesis, which stresses the critical importance of social inputs during the ontogenetic construction of survival-relevant skills. However, the general intelligence explained by the cultural intelligence hypothesis is actually quite limited, so we must seek a more robust explanation for its evolution.

We believe that the cognitive buffer hypothesis (Allman et al. 1993; Deaner et al. 2003; Sol 2009a; 2009b; Lefebvre et al. 2013) offers a better alternative because it can account for phenomena the cultural intelligence hypothesis leaves unexplained. The cognitive buffer hypothesis posits that general intelligence is favored directly by natural selection to help animals cope with novel or unpredictable environments, where it enables individuals to exhibit flexible behavior, and thus find innovative solutions to problems threatening their survival and reproduction. In our view, Burkart et al. dismiss the cognitive buffer hypothesis prematurely. They argue that fundamental preconditions for the evolution of large brains include a slow life history and high survivorship, possible only in species not subject to unavoidable extrinsic mortality such as high predation pressure (van Schaik et al. 2012). However, much can be learned by considering apparent exceptions to "rules" like these, so we offer the octopus as one such exception.

Most octopuses are strictly solitary except when copulating, have very short lives, have countless predators, and produce thousands of offspring, most of which die. Nevertheless, they have some of the largest brains known among invertebrates (Hochner et al. 2006; Zullo & Hochner 2011); they exhibit a great deal of curiosity about their environments (Montgomery 2015); they recognize individual humans (Anderson et al. 2010); they exhibit pronounced individual differences (Sinn et al. 2001; Mather et al. 2012); they use tools; and they play (Mather 1994; Mather & Anderson 1999). Octopuses thus appear to exhibit a considerable amount of general intelligence without any opportunity whatsoever for social learning. Clearly, the cultural intelligence hypothesis cannot account for the general intelligence apparent in creatures like these.

Similarly, the cultural intelligence hypothesis offers little promise with respect to evolving general intelligence in machines. Computer scientists and robotic engineers have understood for decades that the embodiment of intelligent machines affects their ability to adapt and learn via feedback obtained during their interactions with the environment, mediated by sensors and activators (Brooks 1990; 1991; Sharkey & Ziemke 1998; Goldman & de Vignemont 2009). Most hypotheses forwarded to explain the evolution of intelligence in animals, including the cultural intelligence hypothesis, fail to address the question of how morphological traits outside of the nervous system might have shaped intelligence. In creatures such as octopuses and primates, mutations affecting nervous system structure or function, which might generate less-stereotyped and more-flexible behavior, are visible to selective forces in the environment because they can be embodied in the limbs. Thus, greater intelligence is likely to evolve in these animals than in those whose interactions with their environments are more highly constrained.

Roboticists have also realized that logic alone cannot generate much intelligent behavior in their machines, and that to achieve better performance, their robots must also want things. The skills discovered by evolutionary algorithms are diverse, and many such skills may occur within a single population of digital organisms, but individual agents are rarely motivated to acquire a large array of skills. As a result, most current evolutionary algorithms produce domain-specific intelligence in machines that

rarely possess more than a small set of skills, and they are thus suited to performing only tasks that demand that particular skill set. Although an intrinsic motivation to explore the environment has been imitated in artificial agents via machine learning (Schmidhuber 1991; Oudeyer et al. 2007), the production of generalist learners within an evolutionary context remains highly problematic (Stanton & Clune 2016).

Any selection pressure that promotes behavioral diversity or flexibility within the organism's lifetime, including the ability to learn from experience, should theoretically result in enhanced general intelligence. Novel or changing environments should select for individuals who can learn as much as possible in their lifetimes, as suggested by the cognitive buffer hypothesis. Indeed, Stanton and Clune (2016) recently developed an evolutionary algorithm that produces agents who explore their environments and acquire as many skills as possible within their lifetimes while also retaining their existing skills. This algorithm encourages evolution to select for curious agents motivated to interact with things in the environment that they do not yet understand, and engage in behaviors they have not yet mastered. This algorithm has two main components: a fitness function that rewards individuals for expressing as many unique behaviors as possible, and an intra-life novelty score that quantifies the types of behaviors rewarded by the algorithm. Agents are also provided with an intra-life novelty compass that indicates which behaviors are considered novel within the environment. The intra-life novelty compass may simply identify and direct agents toward areas of high expected learning because new knowledge often promotes the ability to perform new skills. Aligned with these results, we suggest that the primary value of the cultural intelligence hypothesis is to offer social learning as an intra-life novelty compass, but that this hypothesis provides neither the requisite fitness function nor anything analogous to an intra-life novelty score.

A viable hypothesis explaining the evolution of large brains and general intelligence should be able to account for general intelligence in any species where it is known to occur, and it should be able to predict the conditions under which we can develop machines with general intelligence as well. The cultural intelligence hypothesis simply cannot do these things.

Where is the evidence for general intelligence in nonhuman animals?

doi:10.1017/S0140525X16001667, e206

Ludwig Huber

Messerli Research Institute, University of Veterinary Medicine Vienna, Medical University of Vienna, University of Vienna, 1210 Vienna, Austria.
ludwig.huber@vetmeduni.ac.at **www.vetmeduni.ac.at/messerli**

Abstract: This commentary contrasts evolutionary plausibility with empirical evidence and cognitive continuity with radiation and convergent evolution. So far, neither within-species nor between-species comparisons on the basis of rigorous experimental and species-appropriate tests substantiate the claims made in the target article. Caution is advisable on meta-analytical comparisons that primarily rely on publication frequencies and overgeneralizations (from murids and primates to other nonhuman animals).

In this thought-provoking, highly inspiring article, Burkart et al. explore the possibility of the existence of general intelligence in nonhuman animals. Given the evidence for *g* in humans, it is a reasonable and worthwhile endeavor to look for its existence in other taxa. However, to pursue a psychometric approach to nonhuman intelligence, it is necessary to obtain relevant and reliable data. As the authors themselves admit, evolutionary plausibility does not amount to empirical evidence.

Within-species comparisons. For more than a century, psychometricians have devised IQ tests to measure human intelligence. However, the breadth of test items is quite narrow. The tasks are, for the most part, administered in the same manner, with no or only modest variation of test-taking situation, motivation, or sensory domain (Locurto et al. 2006). For instance, the WAIS-IV (Wechsler et al. 2008) comprises four index scores, focusing on verbal comprehension, perceptual reasoning, working memory, and processing speed. This paper-and-pencil task may be enough to represent major components of human intelligence, but it does not tap the most interesting cognitive abilities in nonhuman animals, especially in the technical and social domains.

A crucial question in the search for the influence of an underlying general mental ability is the rationale behind which tests are included in the test batteries and the reliability of those tests for uncovering cognitive abilities. Tests measure performance, not cognitive abilities per se. A huge number of possible noncognitive factors may influence performance, from anatomical to perceptual and motivational. Therefore, it is important to know which cognitive tasks and which controls are included in the test battery. Human IQ tests are often constructed in the manner of a best-case scenario, in that tasks are included in the final battery only if they correlate positively with other tasks and loaded positively on the first component. That is, the presence of *g* is assumed and tasks chosen that verify its presence (Locurto et al. 2006). Furthermore, human IQ tests are standardized with several hundreds to thousands of people of all age classes. This is not feasible with (most) nonhuman animals.

Between-species comparisons. Large data sets for valid comparisons are only possible if we collect data from different labs. But can we rely on data sampled in different labs, using (slightly) different methods (different stimuli, apparatuses, procedures, etc.) and groups of subjects differing in important features like housing and rearing conditions, individual experiences, age, and sex composition? This is both a practical and a theoretical problem. It would demand an enormous amount of labor, money, space, and other resources to test a large sample of species in one lab. Even if one has access to a zoo or game park, testing the abilities that tap reasoning in nonhuman cognition is a difficult and time-consuming business. Furthermore, if the tasks were designed to tap different response systems, sensory modalities, and motivations, it would be a huge undertaking.

Therefore, the evidence for general intelligence on the interspecific level so far rests on meta-analyses. This strategy is based on the assumption that the frequency of reported observations of complex traits associated with behavioral flexibility is a reflection of that species' intellectual capability. For instance, Reader and Laland (2002) used indices of innovation, tool use, and social learning for their correlations. But is innovation really a direct outcome of a cognitive trait of a species? The relation is vague and the behavioral definitions are rather slippery. Furthermore, most of these meta-analyses rely on observation frequency, which may deviate widely from the experimentally proven existence of a cognitive trait in a species. For instance, reports of true imitation in callithrichids are very rare, but rigorous laboratory tests have proven its existence (Voelkl & Huber 2000; Voelkl & Huber 2007). The same is true with invisible displacement in *Callithrix jacchus* (Mendes & Huber 2004). Tool use may be the best example of the problem with drawing conclusions about species differences in general intelligence based on publication counting. It is an important ability in chimpanzees, New Caledonian crows, and Galápagos woodpecker finches. However, these species have no clear, experimentally proven cognitive superiority over their non-tool-using relatives, bonobos, carrion crows, or tree finches, respectively (Gruber et al. 2010; Herrmann et al. 2010a; Teschke et al. 2011; 2013). This led to the conclusion that habitual tool use is not a clear predictor of general intelligence, not even physical intelligence (Emery & Clayton 2009). Although

it would be unfair to dismiss the meta-analytical studies completely, at least they require substantiation by experimental data collected with similar methods across large samples of species (Healy & Rowe 2007). So far, such experimental comparisons are rare, and if available, they don't support the meta-analytical studies. All four experimental comparisons listed in Table 5 of Burkart et al.'s target article lack clear-cut evidence for *G*.

Reasoning. Burkart et al. claim that "recent studies are consistent with the presence of general intelligence in mammals" (in the Abstract), which is defined as the ability to reason, plan, and think abstractly (Gottfredson 1997). However, the only cited reasoning study outside of rodents (Anderson 1993; Wass et al. 2012) has not found evidence for *g* (Herrmann & Call 2012). The author of this commentary has found evidence for reasoning by exclusion in several human animals (Aust et al. 2008; Huber 2009; O'Hara et al. 2015; 2016), but so far, evidence for *g* in these species is lacking.

Finally, concerning the search for *g* or *G* in nonhuman animals, caution toward overgeneralization is warranted. The few supportive studies in rodents and primates, two taxa that together represent about 20% of mammalian species and only 2% of vertebrates, cannot be generalized to "nonhuman animals." Especially primatologists may be at risk of overemphasizing cognitive continuity between humans and nonhuman animals, instead of seeing radiation of traits outward in all directions (Hodos & Campbell 1969; Shettleworth 2010a). The search for (human-like) general intelligence (based on reasoning) should be compensated by an appreciation of convergent evolution (Emery & Clayton 2004; 2009; Fitch et al. 2010; Güntürkün & Bugnyar 2016).

The false dichotomy of domain-specific versus domain-general cognition

doi:10.1017/S0140525X16001679, e207

Ivo Jacobs and Peter Gärdenfors

Cognitive Science, Department of Philosophy, Lund University, 221 00, Lund, Sweden.

ivo.jacobs@lucs.lu.se **http://www.fil.lu.se/person/IvoJacobs**
peter.gardenfors@lucs.lu.se **http://www.fil.lu.se/person/PeterGardenfors**

Abstract: The qualitative division between domain-general and domain-specific cognition is unsubstantiated. The distinction is instead better viewed as opposites on a gradual scale, which has more explanatory power and fits current empirical evidence better. We also argue that causal cognition may be more general than social learning, which it often involves.

Burkart et al. view domain-specific and domain-general intelligence as qualitatively different categories and then attempt to find plausible evolutionary scenarios. However, viewing intelligence as a scalar trait is more consistent with biological gradualism. Exclusive categories do not exist for the complex continuous interplay between genes and environment resulting in unique individual development and their evolutionary interactions (Laland et al. 2011; Osvath et al. 2014; Ploeger & Galis 2011; West-Eberhard 2003). The authors also do not explain how any cognitive adaptation can be fully independent of brain size and executive functions, as they posit in Figure 3 of the target article. Although it is true that small brains can house many cognitive adaptations with poor executive functions, they must clearly be at least somewhat related. For instance, primary modules can be inhibited or stored in working memory. The tendency of kittens to respond to small moving objects with behaviours from the hunting repertoire is considered an example of a primary module (Table 2 of target article), yet they can wait for the right time to pounce (inhibitory control) and recall where they have last seen objects (working memory).

The problems of the dichotomy can also be illustrated by considering precocial birds such as ducks and chickens, which are born relatively well-developed; they walk, have open eyes, and forage. Their cognitive abilities can therefore be tested soon after hatching. Precocial animals are perfect for distinguishing between primary and secondary modularization because they can be tested with minimal experience. Filial imprinting occurs when a newly hatched precocial chick limits its social behaviour to a particular object. Under normal circumstances, this means that the chick will attend to and follow its mother. It is one of the most extensively described phenomena in ethology (Bolhuis 1991) and is traditionally considered to be the archetype of instinct, so categorizing it as domain specific and modular should be straightforward. Imprinting indeed appears to be a species-wide adaptive specialisation to a predictable situation that is stable across generations, with relatively quick learning in a specific domain following a characteristic ontogeny.

The concepts of primary modules and instinct resemble each other greatly (see Table 1), so we can criticize them on similar grounds – mainly, that they are not truly qualitatively distinct from their supposed polar opposites (Bateson & Curley 2013; Bolhuis 1991). Many aspects of imprinting go beyond instinct or primary modules. It can be considered domain general and may involve secondary modules because it is phylogenetically and ontogenetically canalized; it is learned until automated; and it can have variable contents with individual differences. Moreover, ducklings understand the relational concept of "same or different" based on imprinting (Martinho & Kacelnik 2016), and chicks are born with advanced folk biology, psychology, and physics (Vallortigara 2012a; 2012b). This suggests massive modularity, which according to Burkart et al. is "entirely compatible with the co-existence of domain-general processes and general intelligence" (sect. 1.2.1., para. 4). It is possible that imprinting is more of the one than the other, but according to their view it has to be *either* general *or* specific, which is incompatible with current empirical evidence.

Table 1 (Jacobs & Gärdenfors). *The description of primary modules by Burkart et al. (Table 2 of target article) strongly resembles the nine different meanings of instinct by Bateson and Curley (2013) when rearranged.*

	Primary Modules	Instinct
Etiology	Evolutionary; reflect natural selection for domain-specific cognitive adaptation	Adapted during evolution; genetic – highly heritable; controlled by a specialised neural module
Development	Skill matures, motor practice (experience-expectant)	Present at birth or particular stage of development; develops before function is established
Content of skills	Pre-set, highly predictable	Developmentally robust – well-canalized; not learned; a functional behavioural system
Distribution	Uniformly present in a given species	Shared by all members of species/sex/age group

The theoretical and empirical evidence for an absolute divide between domain-specific and domain-general intelligence is thus poor. One might argue that certain individuals have general intelligence in the sense that they consistently perform well on various tests across domains, but this hardly seems surprising or controversial. A gradual notion of intelligence means its evolution is more plausible – even repeatedly in different clades (Osvath et al. 2014) – than the potential "hard step" of categorically unique general intelligence.

We are also sceptical of Burkart et al.'s focus on cultural intelligence. Social learning has undoubtedly played an important role in the cognitive evolution of many species, but perhaps it is not as central or exclusive as they claim. In fact, they are concerned that socio-cognitive abilities too often yield inconclusive results or are not even included in test batteries.

Causal cognition can arguably overcome the problems of Table 1 in the target article equally well or better than social learning, which in many cases can be considered to be causal. Woodward (2011) distinguished three levels of causal reasoning (see also Gärdenfors 2003); one can learn to shake a branch to cause fruit to fall because of one's own experience shaking branches (egocentric causal learning), observing others shake branches (agent causal learning), or observing the wind shake branches (observation/action causal learning). It is reasonable that these three levels represent an evolutionary order of expansion of causal cognition. This would constitute another argument that the dichotomy between domain-specific and domain-general intelligence is not plausible.

Rather than learning many one-to-one relations, representing a causal network based on individual and social learning can be highly advantageous and at the base of novel causal interventions (Tomasello & Call 1997; Woodward 2011). This sort of causal cognition can be tested empirically in a variety of species (Blaisdell et al. 2006; Jacobs et al. 2015), and may be of the general nature that Burkart et al. are seeking.

The evolution of fluid intelligence meets formative *g*

doi:10.1017/S0140525X16001680, e208

Kristof Kovacs[a] and Andrew R. A. Conway[b]

[a]*Eszterházy Károly University, 3300, Eger, Hungary;* [b]*Claremont Graduate University, Claremont, CA 91768.*

kristof340@gmail.com **andrew.conway@cgu.edu**

Abstract: The argument by Burkart et al. in the target article relates to fluid (not general) intelligence: a domain-general ability involved in complex, novel problem solving, and strongly related to working memory and executive functions. A formative framework, under which the general factor of intelligence is the common consequence, not the common cause of the covariance among tests is more in line with an evolutionary approach.

The authors present a wide-ranging theory of the evolution of intelligence. However, Burkart et al. seem to have confused the general intelligence factor (psychometric *g*) with fluid intelligence (Gf). Psychometric *g* is a statistical way of describing the positive manifold: the phenomenon that ability tests, each with different content, all correlate positively. As such, psychometric *g* is a necessary algebraic consequence of the positive manifold itself (Krijnen 2004). The concept of general intelligence interprets psychometric *g* as a within-individual, domain-general cognitive ability that permeates all human mental activity so that different tests are functionally equivalent in the sense that they all measure this general ability to a varying extent. This is a sufficient, but not necessary, explanation of the positive manifold. Moreover, it is contradicted by evidence from cognitive neuroscience, neuropsychology, and the study of developmental disorders (e.g., Duncan et al. 1995; Vicari et al. 2007; Wang & Bellugi 1994).

Contrary to *g*, fluid intelligence can be meaningfully conceptualized as a domain-general ability involved in complex, novel problem solving – according to its definition, it is "an expression of the level of relationships which an individual can perceive and act upon when he does not have recourse to answers to such complex issues already stored in memory" (Cattell 1971, p. 115.) or "the use of deliberate and controlled mental operations to solve novel problems that cannot be performed automatically" (McGrew 2009, p. 5). In humans, fluid reasoning is usually measured with tests of nonverbal inductive reasoning. Gf shares nearly half of its variance with working memory (Kane et al. 2005; Oberauer et al. 2005), probably because they both tap executive/attentional processes to a large extent (Engle & Kane 2004).

There are reasons that can lead one to think that Gf and *g* are the same: Gf is central to variation in cognitive abilities to the extent that *g* and Gf are statistically near-indistinguishable (Gustafsson 1984; Matzke et al. 2010). Yet general intelligence and fluid reasoning are clearly different constructs (Blair 2006) – and so are the psychometric factors *g* and Gf (Kovacs et al. 2006). Additionally, whereas the neural substrate of fluid intelligence is in the prefrontal and partly in the parietal cortex (Kane & Engle 2002; Kane 2005), it is difficult to localize *g*, as results depend on the actual battery of tests used to extract *g* (Haier et al. 2009). Also, different components of *g* are differently affected by aging or the Flynn effect (the secular increase in IQ), both of which manifest themselves more strongly on nonverbal than verbal tests (Flynn 2007; Horn & Cattell 1967; Trahan et al. 2014).

Verbal cognition itself is crucial from the target article's perspective when interpreting *g*. In humans, *g* is composed of crystalized intelligence (Gc), too: the ability to apply already acquired skills and knowledge, with an emphasis on language – vocabulary, reading comprehension, and verbal reasoning. This does not translate to nonhuman animals, making it very implausible that general factors reflect the same construct across species. The authors' approach to general intelligence, emphasizing problem solving in novel contexts, also in fact reflects fluid intelligence – the central component of *g*, but not the same as *g*. Finally, executive functions are more strongly related to Gf than to other components of *g* (Conway & Kovacs 2013). In fact, given the authors' emphasis on problem solving in novel situations as well as on the role of cognitive flexibility and executive functions, we often had the impression when reading the target article that Burkart et al. in fact discussed fluid intelligence under the term general intelligence.

If *g* does not reflect a unitary domain-general cognitive ability and is not identical to Gf, then how can the general factor of intelligence be conceptualized? Or, more importantly, if mental tests do not all measure the same general intelligence, then why do tests with different content correlate so strongly?

There are two recent explanations of the positive manifold (with corresponding mathematical formulations) that do not propose a psychological equivalent of psychometric *g:* the mutualism model (van der Maas et al. 2006) and process overlap theory (Kovacs & Conway 2016). Mutualism explains the positive manifold with mutually beneficial interactions between cognitive processes during development. Process overlap theory proposes a functional overlap of cognitive processes when people solve mental test items, such that executive/attentional processes are tapped by a large number of different items whereas domain-specific processes are tapped by specific types of tests only.

Both explanations conceptualize intelligence as a set of independent specific abilities and processes. According to the process overlap theory, *g* is an emergent rather than latent property of mental test scores. Technically, this means that *g* is conceptualized as a formative rather than reflective latent variable: the common consequence of the covariance among tests rather than

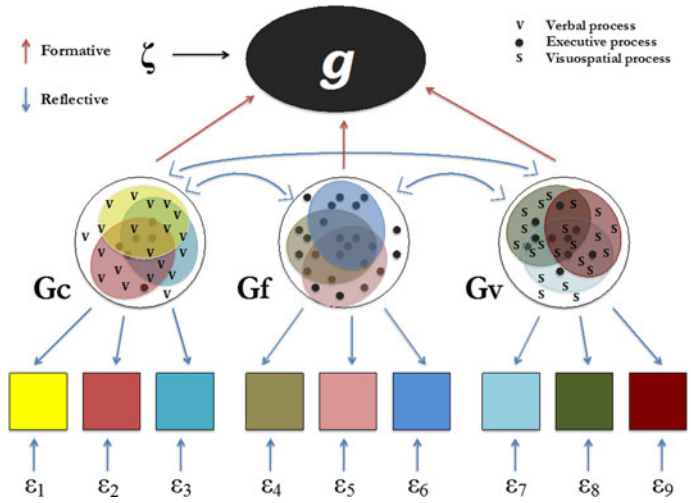


Figure 1 (Kovacs & Conway). The structural model corresponding to process overlap theory on a simplified model.

its common cause. Another common example is socioeconomic status (SES), which clearly is the outcome, and not the cause of a number of indicators like family income, parental education, and so on (Fig. 1).

Such a stance would contribute greatly to the authors' comparative approach, in which *g* would vary from species to species (depending on whether its exact composition includes social skills, language, etc.), whereas a reflective fluid intelligence could indeed be plausibly interpreted as an ability whose evolution was shaped by evolutionary pressures to solve novel problems. The evolution of fluid intelligence could probably be understood through disentangling the evolution of the prefrontal cortex and executive functions in a number of different species.

At the same time, applying a formative framework to *g* could contribute to a functionalist approach, because the primary role of formative constructs is predicting important real-life outcomes (Bagozzi 2007; Howell et al. 2007); in this case, evolutionary ones. Under such a formative/functionalist agenda, the focus would be on individually identifying the cognitive capabilities of each species, ranging from olfactory abilities to social cognition, and how they uniquely contribute to the given species chances of survival and reproduction.

ACKNOWLEDGMENTS

The first author's research was supported by the grant EFOP-3.6.1-16-2016-00001 ("Complex improvement of research capacities and services at Eszterházy Károly University").

Genomic data can illuminate the architecture and evolution of cognitive abilities

doi:10.1017/S0140525X16002004, e209

James J. Lee[a] and Christopher F. Chabris[b,c]

[a]*Department of Psychology, University of Minnesota Twin Cities, Minneapolis, MN 55455;* [b]*Institute for Advanced Application, Geisinger Health System, Lewisburg, PA 17837;* [c]*Department of Psychology, Union College, Schenectady, NY 12308.*

leex2293@umn.edu https://umn.academia.edu/JamesLee
chabris@gmail.com http://www.chabris.com

Abstract: Does general intelligence exist across species, and has it been a target of natural selection? These questions can be addressed with genomic data, which can rule out artifacts by demonstrating that distinct cognitive abilities are genetically correlated and thus share a biological substrate. This work has begun with data from humans and can be extended to other species; it should focus not only on general intelligence but also specific capacities like language and spatial ability.

In 1904, Charles Spearman discovered *g*, the factor measured in common by correlated tests of diverse human mental abilities. The existence of possible *g* homologues in other species and the extent to which the evolutionary trend in our own primate lineage can be characterized as an increase in *g* are among the most important issues facing researchers across the disparate fields interested in cognitive evolution. We applaud Burkart et al. for recognizing the centrality of *g* to any complete understanding of human and animal differences.

Burkart et al. are not alone in expressing concern over the possibility that the correlations between factors defining a statistical *g* (and its between-species analog *G*) might not reflect common information-processing mechanisms (general intelligence) but rather artifacts of various sorts (e.g., Hampshire et al. 2012). In humans, at least, data from twins and genome-wide association studies (GWAS) can rule out various conceivable artifacts by demonstrating that distinct abilities are genetically correlated. The existence of a genetic correlation means that there are polymorphic sites in the genome affecting both traits – either because one trait is on the causal path to the other, or because distinct causal paths emanate toward both traits from a common biological substrate (whose function is influenced by the genetic variants). Empirically estimated genetic correlations between ability tests of different kinds are as large as the simple phenotypic correlations (Kovas & Plomin 2006; Loehlin et al. 2016; Trzaskowski et al. 2013), thus pointing to common biological mechanisms. For instance, Trzaskowski et al. estimated the genetic correlation between *g* and a test of mathematics to be 0.74. A genetic correlation is a coarse-grained summary statistic, but in the near future we believe it will be possible to use DNA-level data to determine whether a given polymorphic site is associated with multiple abilities in a manner consistent with a common mediating mechanism (van der Sluis et al. 2010).

The methodology of GWAS is enabling this revolution because certain special properties of genomic data – such as the natural randomization of genotypes within the offspring of the same parents – enable a high degree of trust in the causal inferences that can be drawn from it (Lee 2012; Lee & Chow 2013; Lee et al. 2016). Unfortunately, genetic methods along these lines may be somewhat difficult to apply to nonhuman species because of the large sample sizes required for adequate statistical power (Chabris et al. 2015). Even in the face of this obstacle,

however, indirect progress may be possible. In recent work, we identified a large number of polymorphic sites in the human genome associated with educational attainment, a heritable trait (Heath et al. 1985) that is genetically correlated with both *g* and intracranial volume (Okbay et al. 2016). More specifically, we found that sites associated with education are much more likely to be found in regions of the genome annotated as likely to affect gene expression in the brain. Armed with such predictive functional annotations, we may be able to determine whether a substitution of one allele for another that has occurred at any point in mammalian evolution would be likely to affect educational attainment – even if the site of the substitution is not polymorphic in modern human populations.

There are some outstanding methodological issues with this approach, such as which parts of the genome should be used as a control for purposes of determining whether likely *g*-affecting sites have undergone an unusual number of base-pair substitutions that would be consistent with the action of natural selection (e.g., Dong et al. 2016). More work also needs to be done to ensure that the functional annotations truly predict causal effects on *g* or some cognitive trait rather than other intermediate phenotypes (e.g., personality traits like neuroticism) that are also genetically correlated with educational attainment in modern Western societies. If these issues can be addressed, however, then many powerful inferences will become possible. For instance, we may be able to find evidence of directional selection increasing *g* in the human lineage or a correlation between the number of substitutions from the time of the common ancestor to the present and the rank of a taxon in some measure of *G* (Johnson et al. 2002). Such findings would bolster many of the points tentatively advanced in the target article, including the identification of the statistical *g*/*G* factors in other species with general intelligence in Homo sapiens.

We also urge Burkart et al. and other researchers to consider important ability factors other than *g*. The correlations between distinct human abilities can be attributed to their common measurement of *g*, but the "error" or "residual" inherent in each ability when it is regarded in this way is also of substantive interest. The authors mention the Cattell-Horn-Carroll (CHC) model of these lower-order factors; setting aside various controversies over terminology and substance, we single out two of these factors because of their ecological validity in the prediction of human behavior (Kell et al. 2013; Lee & Kuncel 2015). The factor that we will call verbal comprehension is characterized by tasks requiring the translation of meanings into verbal units (words, sentences, discourses) and vice versa. Burkart et al. do not emphasize human language, but the search for its evolutionary antecedents has raised many issues – including whether language is independent of other cognitive capacities – that may be illuminated by an interdisciplinary approach (Hauser et al. 2002b; Hurford 2007; Pinker & Jackendoff 2005). We call the other non-*g* factor of interest *spatial visualization*, which is characterized by tasks requiring the mental transformation of representations of objects and scenes in a manner preserving spatial relationships. We suspect a relationship between spatial visualization and tool manufacture analogous to the one between verbal comprehension and language; confirming such a relationship may prove to be a worthwhile research program.

Given the prominence of both language and tool manufacture in human evolution, we are intrigued by the prospect of a mapping between these two capacities and the two arguably most important lower-order ability factors in the hierarchy of human individual differences. Of course, these are not the only abilities relevant to human evolution; various aspects of social cognition, such as face recognition and theory of mind, should also be explored. But in any case, it is now time for this line of research to incorporate and make maximal use of the abundance of genetic data that are becoming available.

Contemporary evolutionary psychology and the evolution of intelligence

doi:10.1017/S0140525X16001692, e210

David M. G. Lewis,[a] Laith Al-Shawaf,[b] and Mike Anderson[a]

[a]*School of Psychology and Exercise Science, Murdoch University, Murdoch, Western Australia 6150;* [b]*Department of Psychology, Bilkent University, 06800 Bilkent, Ankara, Turkey.*

d.lewis@murdoch.edu.au http://davidmglewis.com
laith@bilkent.edu.tr http://laithalshawaf.com
mike.anderson@murdoch.edu.au

Abstract: Burkart et al.'s impressive synthesis will serve as a valuable resource for intelligence research. Despite its strengths, the target article falls short of offering compelling explanations for the evolution of intelligence. Here, we outline its shortcomings, illustrate how these can lead to misguided conclusions about the evolution of intelligence, and suggest ways to address the article's key questions.

Burkart et al. offer an impressive integration of intelligence research across humans and nonhuman species. Their commendable synthesis will serve as a valuable, centralized resource. Despite these strengths, the target article falls short of offering compelling explanations for the evolution of intelligence.

We observe three major issues with the target article. First, it poses multiple questions about intelligence but does not consistently differentiate between them – despite their likely different answers. For example, the question of whether interspecific variation in psychometric intelligence (*G*) exists is fundamentally distinct from whether *G* taps the same construct as within-human variation in intelligence (*g*). Independent of these questions are *why G* exists and *why g* exists – two independent questions that may have different answers.

Conflating these questions can lead to misguided conclusions about the evolution of intelligence. The article establishes the existence of both *G* and *g*. However, it does not logically follow that they therefore (1) tap the same construct or (2) share the same evolutionary origins. First, the authors offer little defense of the implicit position that *g* and *G* tap the same construct. Second, it is plausible that some species exhibit superior performance on intelligence batteries as a consequence of cross-species differences in the information-processing demands of survival- and reproduction-related problems. Individual differences in intelligence among humans may have entirely different origins. Prokosch et al. (2005) proposed that *g* captures individual differences in "developmental stability at the level of brain development and cognitive functioning" (p. 203). For several reasons, this alternative evolutionary model deserves consideration alongside the target article. First, the term "evolved" refers not to the products of just selection, but also of genetic drift, gene flow, and mutation. The target article neglects these non-selective forces and how they could produce *g*. By contrast, Prokosch et al. considered a more comprehensive set of evolutionary forces and posited that *g* reflects the outcome of a balance between selection and genetic mutation. The target article offers no consideration of the mechanistic basis of variation in intelligence. Second, Prokosch and colleagues generated clear, novel predictions based on their model. It is not immediately clear what new predictions the target article's "cultural intelligence" (CI) approach yields. The crucial idea is not that we favor Prokosch et al.'s model, but rather that their work exhibits hallmarks of sound evolutionary science that the CI approach, in its current form, lacks. These include a consideration of selective and non-selective forces, as well as the generation of specific, falsifiable predictions. At present, it is not clear what evidence *could* disconfirm the CI model. We suggest that the CI approach could benefit from more clearly articulating its empirical

predictions, with an emphasis on identifying where it and alternative models advance divergent predictions.

A second, related issue is that the target article attempts to use inappropriate criteria to discriminate between the CI and alternative evolutionary models. Here, we provide four examples of this. First, Figure 1b in the target article presents a pattern of cognitive performance expected from domain-general mechanisms in homogeneous developmental conditions. However, this pattern is identical to that expected when selective forces favor domain-specific mechanisms but non-selective forces (e.g., mutation) impair the performance of these mechanisms. Second, the target article acknowledges that intelligence tests are culture-biased. If we recognize this, then we – the creators of these tests – should certainly acknowledge that they could be *species*-biased. Intelligence batteries tap cognitive performance on different tasks. If the computational demands of these tasks align more closely with the computational demands of the adaptive problems faced by some species, then we should expect interspecific variation in performance on these tasks – *G*. As such, the existence of *G* is *not* "particularly difficult to reconcile" (sect. 2.5, para. 5) with domain-specific mechanisms. We agree with the authors that reconciliation between the massive modularity hypothesis and domain-general views of intelligence is needed, but the mere existence of *G* is insufficient for adjudicating between them.

Third, the target article interprets the absence of "empirical evidence … of specialized adaptive behavioral functions to specific modular neural units" (sec. 1.2.1, para. 2) as evidence against domain-specific mechanisms. This reflects a deep misunderstanding of domain-specificity. A domain-specific mechanism is one that has specialized computational functions, not one that has a delimited neural area.

Fourth, the article ascribes an inability to learn to "primary modules" (sec. 1.2.3, para. 2), which it synonymizes with domain-specific mechanisms. Consequently, the authors use learning as an evidentiary criterion against domain-specific mechanisms. This misconception has been addressed in two recent publications in the flagship journal of the American Psychological Association (e.g., Confer et al. 2010; Lewis et al. [2017]).

These problems point toward our third major issue: the target article badly mischaracterizes contemporary evolutionary psychological thinking. The domain-specific mechanisms proposed by evolutionary psychologists process inputs from the environment, execute computational procedures on these inputs, and produce outputs – *including social learning* (see Henrich & Gil-White 2001; Lewis et al. [2017]). Accordingly, portraying social learning and domain-specific mechanisms as competing alternatives is highly misleading. Domain-specific adaptations can ontogenetically canalize social learning (e.g., see Henrich & Gil-White 2001; see also Karmiloff-Smith's "domain relevant" approach [2015, p. 91]). Crucially, this view squares with the literature presented in the target article *without* forcing the unnecessary and outdated dichotomy between innate versus learned.

We have critiqued several aspects of this article, but we believe it has the potential to advance research on the evolution of intelligence. In particular, the article implicitly points toward cost-benefit analysis as a valuable tool. Applying this tool to cross-species differences in the computational complexity of survival- and reproduction-related problems could be fruitful for understanding *G*. For example, whether a species faces a heterogeneous or homogeneous environment and whether the adaptive problems it faces are characterized by social contingencies (e.g., the psychology of conspecifics) may influence the information-processing complexity of the species' adaptive problems. Comparative analysis of the information-processing complexity of these problems, in conjunction with cost-benefit analyses of the cognitive architecture needed to solve them, has the potential to yield new and testable hypotheses about the evolution of *G*.

G and *g*: Two markers of a general cognitive ability, or none?

doi:10.1017/S0140525X16001709, e211

Charles Locurto

College of the Holy Cross, Worcester, MA 01610.
clocurto@holycross.edu

Abstract: The search for general processes that underlie intelligence in nonhumans has followed two strategies: one that concerns observing differences between nonhuman species (*G*), the second that concerns observing individual differences within a nonhuman species (*g*). This commentary takes issue with both attempts to mark a general factor: Differential responding to contextual variables compromises the search for *G*, and the lack of predictive validity compromises *g*.

The target article by Burkart et al. is a valuable study, bringing together lines of evidence that have heretofore seldom been considered together (Locurto 1997). I do have several concerns about the viability of marking a general factor in nonhumans using either species differences or individual difference. I also have a more minor quibble about the definition of general intelligence (*g*) itself. The authors, quoting Gottfredson (1997, p. 13) offer a rather complex definition of general intelligence that one might call unnecessarily impenetrable, as follows: "the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience" The authors add: "It is thus not merely book learning, a narrow academic skill, or test-taking smarts. Rather it reflects a broader and deeper capability for comprehending our surroundings – 'catching on,' 'making sense' of things, or 'figuring out what to do'" (sect. 1.1, para. 1; from Gottfredson 1997, p. 13).

I offer a simpler definition, based on Charles Spearman's original work in this area. It was Spearman (1904) who first developed the idea of a general faculty, based on his study of individual differences in the performance of school children across a variety of tasks, some sensory/perceptual, as in pitch discrimination, others more fully representative of cognitive functioning, such as school grades (see, in particular, Spearman 1904, p. 291). Spearman defined the general factor as tapping "the eduction of correlates" (or, more fully, "the eduction of relations and correlates," Spearman 1927, pp. 165–66). I love the simplicity and sheer elegance of "the eduction of correlates" expression, and I think it suffices in the stead of more complex definitions. The implication of Spearman's definition was that *g* was better conceptualized as a single process – mental energy and the like – instead of a series of strung-together mechanisms that functioned as a whole because of overlapping microprocesses (see Mackintosh 1998, for presentation of the overlapping mechanisms idea for *g*.). Although the essence of the target article favors density in the definition of *g*, I think Spearman's original simplicity remains defensible.

The marking of a general factor by looking for systematic differences between nonhuman species (*G*) is potentially compromised by Euan Macphail's argument that species differences in cognitive performance may be the result of differences in what he called *contextual variables* (Macphail 1982; 1987) – that is, all of the sensory/motoric/motivational and so on factors that might differ between species, and consequently might masquerade as cognitive differences. The end point of this argument is that we may not be able to reject Macphail's hypothesis that all nonhuman species are capable of all types of learning/cognition. This argument may appear easily rendered moot (after all, isn't a chimpanzee capable of more complex cognition than a frog?), but it has proven more resilient than initially expected. To their credit, the authors cite Macphail's argument, and they offer a reasonable rebuttal in the form that perhaps not all tasks are affected by this problem to the same extent. Reversal learning tasks, for instance, adapt each species to the task in the form of initial acquisition before measuring the rapidity of reversal. Therefore, tasks like this might be seen as mitigating what might be initial between-species differences in

reaction to contextual variables. But the problem posed by contextual variables is more insidious than the authors recognize. To fully account for the influence of these contextual confounds, one would have to expose different species to rather strenuous parametric work, where potential contextual confounds are systematically examined across a given dimension, such as studying species differences in reversal learning across a number of sensory dimensions: visual, olfactory, tactile, and so forth. That kind of work is unlikely to be done, and, as a consequence, Macphail's argument remains a thorn in our collective side.

The study of within-species individual differences is a more promising avenue for identifying markers of a general process. Systematic individual differences have been observed in nonhumans, particularly in mice, and these differences are not confounded by differences in noncognitive factors: for instance, overall activity levels (Locurto & Scanlon 1998; Locurto et al. 2006). However, an important, perhaps even critical limitation of such studies is that they lack something that is commonplace in studies of human *g* – namely, what is called *predictive* or *criterion-related validity* (Anastasi 1961). In psychometrics, validity refers to what a test measures. Predictive validity refers to the effectiveness of a test in forecasting behavior in domains outside of the test content per se. To assess it, there need to be independent measures of what the test is designed to predict. Independence in this sense can be taken to mean measures outside of the province of the test items themselves. In the human literature, predictive validity of an intelligence test is not at issue: *g* is a reasonably good predictor of various measures of life outcome, including school achievement, the probability of occupational success, social mobility, and even health and survival. *g* is better at predicting such variables than are specific cognitive abilities on their own (Locurto 1991). The many criteria external to the test itself that correlate with human *g* represent a powerful measure of real-life success.

There is nothing similar in the nonhuman literature on *g*, although there have been important findings that stretch the initial *g* battery to include a number of additional processes that seem reasonably related to what *g* should measure, such as selective attention, working memory, and tests of reasoning (Matzel et al. 2011b; Sauce et al. 2014). These extensions are valuable, but they do not constitute extra-domain assays. They are simply additional cognitive tasks that load on the initial *g*. This form of adding tasks is itself a type of validity called *content validity*, but it is not predictive validity. The authors recognize this issue, and in their Table 7 they offer a series of additional categories of evidence, some of which are forms of predictive validity, that would be useful going forward. The authors end by raising the critical question: does (nonhuman) *g* predict success in real life? Only if that question can be successfully addressed can we conclude that *g* is not uniquely human.

Hierarchy, multidomain modules, and the evolution of intelligence

doi:10.1017/S0140525X16001710, e212

Mauricio de Jesus Dias Martins[a,b] and Laura Desirèe Di Paolo[c,d]

[a]*Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany;* [b]*Humboldt Universität zu Berlin, Berlin School of Mind and Brain, 10117 Berlin, Germany;* [c]*Lichtenberg-Kolleg Institute, University of Goettingen, 37075 Göttingen Germany;* [d]*Leibniz ScienceCampus and Research Group Primate Social Evolution DPZ-German Primate Centre, 37077 Göttingen, Germany*

mmartins@cbs.mpg.de

https://www.researchgate.net/profile/Mauricio_Martins4

lauradesiree.dipaolo@gmail.com

https://sites.google.com/site/lauradesireedipaolo

Abstract: In this commentary, we support a complex, mosaic, and multimodal approach to the evolution of intelligence. Using the arcuate fasciculus as an example of discontinuity in the evolution of neurobiological architectures, we argue that the strict dichotomy of modules versus *G*, adopted by Burkart et al. in the target article, is insufficient to interpret the available statistical and experimental evidence.

Burkart et al.'s premise is that cognitive abilities can be supported either by the evolution of "primary modules" (sect. 1.2.3, para. 2; domain-specific adaptations to specific environmental challenges), or by the expansion of domain-general intelligence (G). If this premise were true, then the current empirical research, based largely on principal component analyses (PCAs), would be more consistent with the idea that a large portion of cognition in several species is explained by G rather than by collections of primary modules. Reviewing this empirical literature, the authors admit the results are somewhat ambiguous. Nevertheless, they predict we will find stronger evidence for the evolution of G in the future, because the data seem largely inconsistent with the primary modular perspective.

Here, we argue that the strict dichotomy of primary module versus G is misleading: There are occasional evolutionary *discontinuities* in neurobiological architectures that support a range of cognitive abilities, which are neither domain general nor *modular adaptations* for specific environmental challenges.

Our target example is the arcuate fasciculus (AF), which is a neural fiber tract enabling a direct connection between temporal cortex (including auditory cortex) and inferior frontal gyrus (involved in cognitive control) (Catani et al. 2005). This tract, exceptionally well developed in humans in comparison with other primates (Rilling et al. 2008), is a neurobiological evolutionary discontinuity. By neurally binding the regions responsible for auditory processing and cognitive control, this new architectural feature greatly enhanced (1) the working memory for verbal information (vWM) – quite poor in nonhuman primates (Plakke et al. 2015; Scott et al. 2012); and (2) the capacity to process sequences (Dehaene et al. 2015).

This peculiar connectivity pattern seems to be a crucial prerequisite for the evolution of multiple abilities relying on hierarchical *sequential* structure (e.g., language, music, and complex action) (Fadiga et al. 2009; Fitch & Martins 2014). However, improvements in vWM and sequence processing do not necessary permeate other (nonsequential) cognitive domains, thus not allowing any interpretation in terms of modules or *G*. For instance: (1) Some nonhuman primates (e.g., chimpanzees) seem to show spatial WM superior to that of humans (Inoue & Matsuzawa 2007), and (2) although the capacity to represent social hierarchies seems to be within the range of nonhuman primate cognition (Seyfarth & Cheney 2014), and the ability to process spatial hierarchies is conserved among nonhuman mammals (Geva-Sagiv et al. 2015), the capacity to process sequential structures nonetheless remains limited in these clades.

Another source of evidence for this specialization comes from neuroimaging. Although the processing of sequential hierarchies activates the inferior frontal gyrus (a region strongly connected with the AF) (Fadiga et al. 2009; Fitch & Martins 2014), the same is not true for nonsequential hierarchies in the visual, spatial, and social domains (Aminoff et al. 2007; Kumaran et al. 2012; Martins et al. 2014). Instead, the latter group of hierarchies seems to be represented by a domain-general episodic memory system.

This cognitive mosaic argues against a simple gradual expansion of *G*. When performing a PCA, including individuals of different primate species, the emergence of the human AF (and enhanced vWM) would be more easily classified as multidomain or multipurpose cognitive ability, but neither as domain-specific (because it increases the capacity within a range of domains) nor as domain general (because these improvements are specific to sequential but not to nonsequential domains).

In sum, we suspect that the research program advanced by Burkart et al. is designed to distinguish only between modules and G, leaving aside other possible interpretations that would fit better with the available data (e.g., Anderson 2016; Karmiloff-

Smith 2015). In our opinion, a third way between modules and G will give a more suitable account for clade-specific discontinuities (grounded on neurobiological architectural changes), which would fit better the statistical models. These discontinuities offer a great opportunity to capture capacities that are neither gradual expansions of G nor specific modular adaptations to specific environmental problems. Therefore, they are required to overcome intrinsic limitations of current models, theoretically improving them and achieving a more realistic account of the evolution of cognition across different species.

Evolution, brain size, and variations in intelligence

doi:10.1017/S0140525X16001722, e213

Louis D. Matzel and Bruno Sauce

Department of Psychology, Program in Behavioral and Systems Neuroscience, Rutgers University, Piscataway, NJ 08854.

matzel@rci.rutgers.edu

https://www.researchgate.net/profile/Louis_Matzel

sauce.bruno@rutgers.edu

https://www.researchgate.net/profile/Bruno_Sauce

Abstract: Across taxonomic subfamilies, variations in intelligence (*G*) are sometimes related to brain size. However, within species, brain size plays a smaller role in explaining variations in general intelligence (*g*), and the cause-and-effect relationship may be opposite to what appears intuitive. Instead, individual differences in intelligence may reflect variations in domain-general processes that are only superficially related to brain size.

The "evolution" of interest in individual differences in the intelligence of nonhuman animals has followed a circuitous route. Individual differences in intelligence were a central focus of early psychologists (note the inclusion of chapters in our first textbooks; e.g., Seashore 1923), as well as, some decades later, of the first animal learning theorists (e.g., Thorndike's studies in the 1930s). However, with the increasing fixation on the "experimental approach" and reductionism, interest in individual differences waned overall, and systematic studies of variations in intelligence *within* animal species were virtually abandoned between 1940 and 2000. This trend has shifted dramatically in recent times, with increasing interest in between-species comparisons of intelligence (*G*), and more dramatically, in within-species variations in intelligence (*g*). In this spirit, Burkart et al. have done commendable work summarizing the advances, insights, and limitations of animal research on individual differences in intelligence, and have placed this work in the important context of contemporary evolution theory.

Although we agree with many of Burkart et al.'s conclusions, we are skeptical of their inference that the evolution of intelligence, as well as individual differences in intelligence, is inextricably tied to brain size. Brain size does appear to explain differences in the cognitive capacities of closely related species, although the relationship begins to break down across families and higher taxonomic groups. It is similarly problematic that while Neanderthal brain size ranged from 1,300–1,600 grams, their human counterparts had brain sizes of 1,200–1,500 grams. Current theory suggests that competition between the cognitively superior humans and cognitively inferior Neanderthal accounted for the latter's rapid extinction (Banks et al. 2008; Gilpin et al. 2016). Relatedly, the size of the human brain has *decreased* during the last 100,000 years (Aiello & Dean 1990), a time during which we underwent unusually rapid cognitive gains.

Although brain size does have *some* value in explaining the cognitive capacities of closely related species (i.e., *G*), it is less successful when applied to individual differences *within* a species. Early estimates suggested a weak relationship between brain size and intelligence ($r^2 = 0.02$–0.07; reviewed in Van Valen 1974), and meta-analyses based on modern imaging techniques find only a marginal increase in this estimate ($r^2 = 0.08$; reviewed in McDaniel 2005). Furthermore, the strength of correlations between brain size and intelligence vary across specialized abilities, and in the case of some abilities, *no* correlation is observed (van Leeuwen et al. 2009; Wickett et al. 2000), suggesting that variations in brain size may instantiate differences in specific abilities, but not variations in *general* intelligence. So why might any correlation exist between brain size and intelligence? A possibility that is widely ignored is that more intelligent individuals interact more extensively with their environments (e.g., they explore more, they learn more; Light et al. 2011; Matzel et al. 2006), and this "environmental enrichment" promotes brain growth (Rosenzweig & Bennett 1996). Simply stated, brain size might be *influenced* by intelligence, but might not itself *cause* differences in intelligence. This possibility has received wide support outside of the field of intelligence (Clayton 2001; Maguire et al. 2000; van Praag et al. 2000; Will et al. 2004), and can explain the paradoxical observation that the correlation between IQ and brain size only emerges after age 7 (by which time differential experiences will have begun to accumulate; McDaniel 2005).

The role of brain size in intelligence may matter less than we intuit. It is important to be reminded that brain size is only a very indirect measure of how general intelligence is instantiated. Higher cognition is highly complex, and the circuitry, neurochemistry, and intracellular components of the brain all contribute to its computational capacity. For example, as noted by Burkart et al., we have reported that general intelligence in mice is correlated with the expression in the prefrontal cortex (PFC) of a dopaminergic gene cluster(Kolata et al. 2010), and smarter mice express higher dopamine-induced activity in the prefrontal cortex (Wass et al. 2013). In humans, the dopaminergic system in the PFC seems also to be closely linked with executive functions and intelligence (McNab et al. 2009; Miller & Cohen 2001). And whereas the brain of birds differs strikingly from the mammalian brain (e.g., it lacks the 6 layers of lamination in the neocortex), the avian nidopallium caudolaterale (NCL) is remarkably similar to the mammalian PFC. Like the PFC, the NCL is a hub of multimodal integration connecting the higher-order sensory input to limbic and motor structures (Gunturkun & Kroner 1999) , and dopamine in the avian NCL seems to play a similar functional role in higher cognition as it does in the mammalian PFC (Karakuyu et al. 2007; Veit et al. 2014). This confluence of evidence across taxonomic groups (humans, mice, and birds) is compelling, and at least as parsimonious as the descriptions of intelligence based on variations in brain size.

Burkart et al. imply in their current article and state explicitly elsewhere (van Schaik et al. 2012) that "general intelligence is not a uniquely derived human trait but instead a phylogenetically old phenomenon, found among primates, rodents and birds" (p. 280). However, the PFC and NCL are on opposite ends of the cerebrum and possess distinct genetic expression patterns, leading some to claim that these regions are not homologous but, rather, represent a case of evolutionary convergence (Gunturkun 2012). Thus, non-homologous fields converged over the course of 300 million years into mammalian and avian prefrontal areas that generate the same cognitive functions (e.g., working memory capacity; Diekamp et al. 2002; Matzel et al. 2013) that contribute to the establishment of general intelligence. In other words, general intelligence could have evolved multiple times in different taxonomic groups. Of course this is a matter of considerable controversy (Karten 2015), and the question is far from resolved. Nonetheless, this type of solution is more parsimonious than one based solely on brain size, and mitigates the extant problem of the "cost" of bigger brains. We hope that the "evolution" of interest in the variation in general intelligence follows this route for the next decade.

Habit formation generates secondary modules that emulate the efficiency of evolved behavior

doi:10.1017/S0140525X16001734, e214

Samuel A. Nordli and Peter M. Todd

Department of Psychological and Brain Sciences and the Cognitive Science Program, Indiana University, Bloomington, IN 47405-7007.

snordli@indiana.edu pmtodd@indiana.edu

http://www.indiana.edu/~abcwest/

Abstract: We discuss the evolutionary implications of connections drawn between the authors' learned "secondary modules" and the habit-formation system that appears to be ubiquitous among vertebrates. Prior to any subsequent coevolution with social learning, we suggest that aspects of general intelligence likely arose in tandem with mechanisms of adaptive motor control that rely on basal ganglia circuitry.

Burkart et al. conclude that many or all vertebrate minds consist of two sets of modular skills: one hardwired and the other ontogenetically constructed – primary and secondary skills respectively. They seem to define general intelligence operationally in terms of its facilitatory role in the process by which an organism obtains secondary skills. If this is correct, the next step should be to identify and understand the neurobiological mechanisms that underlie this process. Studying the evolution of general intelligence in terms of the evolutionary history of its constituent neural structures should provide valuable direction to the comparative research recommended by Burkart et al. Here, we offer a mechanistic explanation for these two skill types and why they could be indistinguishable from one another; this centers around the basal ganglia, a set of subcortical nuclei that are good candidate structures for the foundations of general intelligence. In particular, learning and automatizing secondary skills appears to be what cognitive neuroscientists describe as habit formation, a process reliant on basal ganglia circuitry (Smith & Graybiel 2016).

Certain aspects of almost any ecological context cannot be directly addressed by natural selection simply because those aspects are too transitory to support intergenerational selection, which is required for specific adaptive solutions to evolve. Although inaccessible to evolved primary skills, some transitory ecological factors nevertheless remain stable for significant portions of an individual organism's life span – the specific geographical features surrounding a given animal's den, for example, or the physical layout of the controls in a personal automobile (and the associated actions required to operate those controls). Transitory stability is itself an ecological factor that is susceptible to adaptive exploitation via natural selection. Species that evolved the general capacity to identify, learn, and exploit instances of transitory ecological stability will have obtained organismal objectives more successfully and more efficiently than species lacking this capacity. Secondary skills as described by Burkart et al. are the exploitative products of such a capacity; that is, although evolved primary skills exploit the stability of specific perennial factors, the capacity to form secondary skills exploits transitory stability as a perennially general characteristic of temporary factors (Nordli 2012). From this perspective, as primary and secondary skills have each arisen to exploit ecological stability (whether long-term or transitory), it is unsurprising that their shared properties – speed, efficiency, automaticity – render them effectively indistinguishable (absent knowledge of their ontogeny), as Burkart et al. point out.

Supporting this perspective, research suggests that primary and secondary skills are each encoded within basal ganglia-based circuits (Graybiel 1995), such as the fixed sequence of grooming behaviors that is ubiquitously exhibited in rats (Aldridge et al. 2004) and learned paths rats take in a maze (Barnes et al. 2005). These nuclei also appear to be integral to goal-directed action selection, stringing behaviors together in service of achieving contextualized reinforcement (Graybiel 2008). As a sequence of reinforced behaviors (e.g., a maze path) is repeated within a specific context (e.g., a chocolate reward at the end), the entire string is encoded within the basal ganglia as a single behavioral "chunk" that then exhibits quick and efficient cue-based automaticity (Jin et al. 2014). The basal ganglia contribute centrally to cognition as well, through working memory, attention, decision making, and other processes (Stocco et al. 2010). In this context, it is intriguing that mice with a humanized version of their Foxp2 gene – the so-called "language gene" – develop neurons in the basal ganglia with increased plasticity and significantly lengthened dendrites (Enard et al. 2009), and also exhibit accelerated habit learning relative to normal mice (Schreiweis et al. 2014).

If general intelligence is the set of processes that allow organisms to discover, learn, and automatize secondary skills, the basal ganglia may be largely responsible for much of what qualifies as intelligence: (1) modulating rewards to direct or "canalize" attention and motivate goals (e.g., inclining to attend toward and imitate conspecifics); (2) exploring a potential action space and achieving targeted objectives by selecting goal-directed behaviors (e.g., practicing/refining an approximation of what conspecifics do); and (3) automatizing contextualized behavioral sequences that have been repeatedly reinforced (e.g., reproducing efficient skill behavior that may now be imitated by others). This is consistent with the cultural intelligence perspective advocated by Burkart et al., but the coevolutionary enhancement of social learning and general intelligence does not itself explain the evolutionary origins of general intelligence. Instead, we should look to basal ganglia circuitry, which is functionally conserved across all vertebrate species, and which likely evolved over 560 million years ago (Reiner 2010; Stephenson-Jones et al. 2011).

The most basic function of basal ganglia circuitry is adaptive motor control, directing goal-oriented motor sequences (Grillner et al. 2013). Energy demands and reproductive success are the main fitness pressures, and most organisms adapt to that pressure by moving about through space – foraging for food and mates – as efficiently and effectively as possible (see Stephens & Krebs 1986); competition over limited resources in these domains likely resulted in an evolutionary arms race. The capacity to automatize stereotyped patterns of learned motor behaviors into secondary skills is a powerful weapon in that war, enabling the execution of learned motor patterns with the same speed, efficiency, and specificity of evolved motor patterns. We suspect that general intelligence initially coevolved with mechanisms of adaptive motor control to facilitate the search for and learning of new adaptive motor skills. This intelligent search capacity may subsequently have been generalized through exaptation to facilitate the search for new adaptive cognitive skills: spatial foraging and searching through memory space appear to be expressions of the same general exploratory capacity, the goal-directed nature of which is modulated by basal ganglia circuitry via the dopaminergic reward system (Hills et al. 2008; Hills et al. 2015). Comparative investigations of interspecies differences in basal ganglia-based circuitry may provide further clues regarding the evolution of general intelligence; future studies should pay attention to these structures and the mechanisms of habit formation to which they contribute.

The evolution of analytic thought?

doi:10.1017/S0140525X16001746, e215

Gordon Pennycook and David G. Rand

Department of Psychology, Yale University, New Haven, CT 06520-8205.

gordon.pennycook@yale.edu https://gordonpennycook.net

david.rand@yale.edu davidrand-cooperation.com

Abstract: We argue that the truly unique aspect of human intelligence is not the variety of cognitive skills that are ontogenetically constructed, but

rather the capacity to *decide* when to develop and apply said skills. Even if there is good evidence for *g* in nonhuman animals, we are left with major questions about how the disposition to think analytically can evolve.

In their admirably broad discussion, Burkart et al. review many important distinctions in the study of human cognition, including fluid versus crystallized intelligence and domain-general versus domain-specific mechanisms. Nonetheless, by focusing on *g*, the authors did not acknowledge that individual aspects of human intelligence – some of which presumably evolved separately – may have been particularly important for the evolution of human intelligence. In our view, the capacity to *decide* when to develop and use intellectual skills is not only a crucial aspect of human intelligence, but also it may in fact be unique to *human* intelligence. Human metacognition of this sort was not discussed by Burkart et al.

Consider the following problem (Frederick 2005):

A bat and ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?

Most educated adult humans are capable of generating a response to this question intuitively and automatically (namely: 10 cents). This likely occurs through a domain-general canalization process (see "The urgency problem" in the target article, Table 1). However, the automatic response is not the correct answer (if the ball cost 10 cents, the bat would have to cost $1.10 and together they would cost $1.20 – the correct answer is 5 cents). Although the majority of people simply give the incorrect intuitive response to this question (e.g., 64.9% of University of Waterloo undergraduates; Pennycook et al. 2016a), some are capable of answering it correctly. This exercise of intelligence requires not just the capacity to solve the problem, but also the willingness to apply effortful cognitive processing to a problem despite the presence of what initially appears to be a suitable response (Stanovich & West 1998; 2000). There is now a great deal of evidence that human rationality (however imperfect, see Kahneman 2011; Kahneman & Frederick 2005) involves not simply computational cognitive operations (i.e., *g*), but also algorithmic-level operations that determine the course of reasoning and decision making (see Stanovich 2009a; 2009b; 2011).

Moreover, recent research indicates that the propensity to think analytically as a means to override automatic responses has consequences for our everyday lives (Pennycook et al. 2015b). For example, more analytic individuals have less-traditional moral values (Pennycook et al. 2014; Royzman et al. 2014) and are less likely to hold beliefs that are religious (Gervais & Norenzayan 2012; Pennycook et al. 2012; 2016b; Shenhav et al. 2012), paranormal (Pennycook et al. 2012), and/or conspiratorial (Swami et al. 2014). Analytic thinking disposition has also been linked with increased acceptance of science (Gervais 2015; Shtulman & McCallum 2014) and lowered acceptance of complementary and alternative medicine (Browne et al. 2015) and pseudo-profound bullshit (Pennycook et al. 2015a). Analytic thinking can also undermine cooperation and prosociality (Rand 2016; Rand et al. 2016; Rand et al. 2014; Rand et al. 2012), as well as punishment (Grimm & Mengel 2011; Halali et al. 2014; Sutter et al. 2003).

Consideration of the evolutionary dynamics of metacognition is, therefore, of key importance for understanding the evolution of human intelligence (Bear & Rand 2016b). Recent work using formal evolutionary game theory models has begun to shed light on this issue from a theoretical perspective, both in the domains of intertemporal choice (Tomlin et al. 2015; Toupo et al. 2015) and cooperation (Bear & Rand 2016a; Bear et al. 2016). These models illustrate how the willingness to override intuitive responses can be favored by natural selection in settings where flexibility and planning are particularly useful, and also how complex cyclical dynamics of automatic versus controlled cognition can emerge. This growing body of theoretical work calls for empirical examination of cognitive control in nonhuman animals (e.g., MacLean et al. 2014; Rosati & Santos 2016).

Burkart et al. discuss executive functions like inhibitory control, working memory, and cognitive flexibility (sect. 1.1) and highlight the importance of "reasoning ability and behavioral flexibility" for human and nonhuman intelligence (sect. 1.1, para. 1). Thus, the human capacity for overriding intuitive outputs (such as 10 cents in the bat-and-ball problem) is clearly acknowledged. Nonetheless, treating these aspects of human cognition as other types of cognitive processes suppresses a distinction we think should be emphasized. Can humans alone *decide* when (or if) to initiate cognitive processes, as well as when (or if) to reflect upon their outputs? The findings highlighted previously suggest that the capacity to *decide to think* is a core intellectual skill that distinguishes humans from each other. We assert that this skill is also crucial to distinguishing humans from nonhuman animals.

Although we agree that the pursuit of *g* (and *G*) in nonhuman animals is worthwhile, it is not simply that the current body of work is preliminary (as the authors state). Rather, understanding the evolution of human intelligence requires a broader view of human rationality. Thus, unfortunately, we are even further from definitive conclusions than is intimated by the target article. Even if there is good evidence for *g* in nonhuman animals and this ultimately informs us about the evolution of cognitive skills in humans, we will still be left with major questions about how the human capacity to decide when to think (i.e., the disposition to think analytically, over and above *g*) can evolve.

"Birdbrains" should not be ignored in studying the evolution of *g*

doi:10.1017/S0140525X16001758, e216

Irene M. Pepperberg

Department of Psychology, Harvard University, Cambridge, MA 02138.
impepper@media.mit.edu **impepper@wjh.harvard.edu**

Abstract: The authors evaluate evidence for general intelligence (*g*) in nonhumans but lean heavily toward mammalian data. They mention, but do not discuss in detail, evidence for *g* in nonmammalian species, for which substantive material exists. I refer to a number of avian studies, particularly in corvids and parrots, which would add breadth to the material presented in the target article.

I agree with the authors' basic thesis, depicted in the target article's Figure 3, which argues for combinations of heritable and learned abilities that result in general intellectual achievements, whether in humans or nonhumans. However, in their discussion of nonhuman subjects, the emphasis on nonhuman primate, and to a lesser extent mammalian, species is discouraging. The few references to avian cognition do not do justice to the wide variety of abilities and studies – sometimes in a single individual or species – that provide evidence for generalized intelligence. Thus, the point of my commentary is to advocate strongly for avian *g*.

I do, of course, understand that formal *g* analyses (i.e., batteries of different tests on numerous individuals within and across species) are lacking for avian cognitive capacities, as compared to the several existent analyses on nonhuman primates and rodents. Nevertheless, I hoped that Burkart et al. would have discussed the large number of studies on a wide variety of topics performed on avian species, particularly on corvids and psittacids … and maybe would have attempted some kind of review, if not analysis, of their own. For example, early in the article the authors suggest that *transfer* of knowledge from one domain to another novel context provides evidence for *g*, yet little discussion exists of instances of such behavior in avian subjects (for example, transfer of the trained use of the label "none" from describing the absence of similarity and difference of specific attributes between objects [Pepperberg 1988] to the spontaneous use of the label for describing the absence of a size differential

[Pepperberg & Brezinsky 1991] to the additional spontaneous use for describing the absence of a specific numerical set in a collection [a zero-like concept; Pepperberg & Gordon 2005]).

The authors note that *g* has "robust correlates to brain structure and function" (sect. 1.1, para. 3), yet do not mention significant research on avian brains. Recent studies, although not always performing the correlations themselves, have shown that avian brain anatomy would correlate with many *g*-related abilities; see, for example, in addition to those references cited, Iwaniuk et al. (2009) or Jarvis et al. (2005). Specifically, Olkowicz et al. (2016) have found that parrots and corvids have forebrain neuron counts equal to or greater than primates with much larger sized brains, and argue that this finding likely explains the advanced cognitive abilities found in these avian species.

In general, the authors only partially consider parallel/convergent evolution of *g* with respect to avian species, again particularly concerning corvids and parrots. For example, K-selected traits (e.g., long lives, few offspring that are relatively slow to mature, etc.), are discussed at various points in the target article as being associated with the evolution of *g*; these traits are found in most corvid and parrot species. The authors highlight the importance of sociality in driving intelligent behavior (sensu Jolly [1966] and Humphrey [1976]): Notably, the majority of parrot and corvid species live in complex social groups, and evidence exists for several types of learning that are enhanced via conspecific or allospecific social influences (e.g., corvids: Miller et al. 2014; New Zealand kea parrots: Heyse 2012). Myriad papers on corvid social cognition have been published by Bugnyar and his colleagues (e.g., Bugnyar & Heinrich 2006). Acquisition of referential use of human speech by Grey parrots occurs through social learning (e.g., Pepperberg 1981; 1999); similarly, "bilingual" songbirds learn the form and likely use of heterospecific vocalizations via intense social interaction (e.g., Baptista 1981).

Furthermore, as with nonhuman primates, the dominance hierarchies that are prevalent in social groups of corvids (e.g., Chiarati et al. 2010) require an understanding of advanced cognitive processes such as individual recognition (e.g., Izawa & Watanabe 2008) and transitive inference (e.g., Paz-y-Miño et al. 2004). Granted, the references I cite involve different corvid species; nevertheless, the cognitive requirements across species would likely be quite similar given their similar ecology/ethology. And, although dominance hierarchies in parrot flocks have not been studied in the wild, hierarchies have been observed in captivity (Szabo et al. 2016; parrots in my lab also exhibit a hierarchy), and understanding linear ordering can also be related to cognitive capacities such as the spontaneous comprehension of ordinality (Pepperberg 2006).

Tests used as evidence of general intelligence, even by the authors' admission, are mostly basic, but other tests, even if performed on only a limited number of subjects, strongly demonstrate advanced avian capacities. For example, evidence for executive function (planning, delayed gratification) is evident in corvids (Hillemann et al. 2014; Raby et al. 2007). One cannot argue that such behavior is modularly related to caching, as success on the same tasks can be seen in parrots that do not cache (Auersperg et al. 2013; Koepke et al. 2015). Grey parrots understand not only categories (e.g., what is or is not green), but also concepts such as "color," "shape," and "matter" (i.e., the existence of these hierarchical concepts, under which categories such as green and wood are sorted; Pepperberg 1983) and that two objects can be related based on just a subset of these concepts; that is, for second-order concepts of same-different (Pepperberg 1987). The authors mention reasoning by exclusion: for such abilities in parrots and corvids, see Pepperberg et al. (2013), Schloegl (2011), and Schloegl et al. (2009). Likewise, for advanced avian understanding of number concepts, see Smirnova (2013) and Ujfalussy et al. (2014); these abilities are often at a level more advanced than those shown to date for nonhuman primates (e.g., Pepperberg 2006; Pepperberg & Carey 2012). Research papers on tool use by corvids that do not use tools in nature are too numerous to mention; for aspects of physical cognition in parrots, note van Horik and Emery (2016).

The authors have, essentially, performed a meta-analysis on a number of meta-analyses, and I have no arguments about their basic thesis – my criticism is merely that readers interested in this thesis, particularly readers with little knowledge of nonhuman capacities, would unfortunately be left unaware of a large number of striking *avian* abilities that provide considerable evidence for *g*.

General intelligence is an emerging property, not an evolutionary puzzle

doi:10.1017/S0140525X1600176X, e217

Franck Ramus

CNRS, Ecole Normale Supérieure, EHESS, PSL Research University, 75005 Paris, France.

franck.ramus@ens.fr http://www.lscp.net/persons/ramus/en/

Abstract: Burkart et al. contend that general intelligence poses a major evolutionary puzzle. This assertion presupposes a reification of general intelligence – that is, assuming that it is one "thing" that must have been selected as such. However, viewing general intelligence as an emerging property of multiple cognitive abilities (each with their own selective advantage) requires no additional evolutionary explanation.

As the authors acknowledge, the concept of general intelligence is empirically grounded solely in the observation of positive correlations between all test scores, as reflected by a general factor termed *g* explaining a large share of variance in all tests (Spearman 1904). All other accounts are simply debatable interpretations or hypotheses attempting to relate *g* to some other cognitive or biological constructs. They run the risk of reifying what is primarily a statistical construct, and also of seriously confusing the search for an evolutionary explanation. For instance, Gottfredson's (1997) definition of intelligence is little more than a scholarly formulation of the folk concept of intelligence, but offers no guarantee of matching psychometric *g*. Burkart et al. initially conflate *g* with executive functions, but this changes the nature of the problem. If general intelligence reduced to executive functions, then to the extent that each executive function offers a selective advantage, the evolution of general intelligence would not be a major puzzle. Similarly, general intelligence is also identified with domain-general cognitive processes, which is a different, and unnecessary, hypothesis as we will show. Furthermore, many putative domain-general cognitive functions turn out to be less general than they seem. For instance, there are separate working memory systems for verbal, visuospatial, and other modalities. Similarly, words such as *inhibition* and *attention* wrongly suggest unitary phenomena, whereas they are used to describe a host of distinct processes, none of which can be said to be truly domain-general, and none of which is an evolutionary puzzle. Finally, certain cognitive functions can serve domain-general purposes while having been selected for more specific adaptive value. This may be the case of language, which serves as a mediator across many cognitive functions, yet may have evolved for purely communicative purposes (Jackendoff 1999; Pinker & Bloom 1990).

More generally, every attempt to reduce general intelligence to a single cognitive (processing speed, working memory, etc.) or biological (brain volume, nerve conduction velocity, etc.) construct has failed, each construct showing moderate correlation with *g* and being best described as simply one contributor to the *g* factor (e.g., Mackintosh 2011). Thus, trying to tackle the evolution of general intelligence by addressing the evolution of any of these constructs is a form of attribute substitution (Kahneman & Frederick 2002).

Understanding the evolution of psychometric *g* requires understanding how it comes about. As early as 1916, Thomson (1916) showed that it is sufficient to postulate underlying group factors that influence several tests to obtain a positive manifold without a general factor (see also Bartholomew et al. 2009). Reframed in modern psychological terms, an elementary analysis of tests shows that no test is a pure measure of a cognitive function (or construct). The relationship between cognitive functions and test scores is many-to-many: Each test score is influenced by several cognitive functions, and each cognitive function influences several test scores (in the same direction). The latter observation suffices to explain that test scores are positively correlated. We submit that the logic of Thomson's bonds model is much more general, as it also applies to factors underlying cognitive functions. Indeed, each brain function or property (e.g., frontal gray matter volume, nerve conductance velocity, dopamine synthesis, etc.) influences several cognitive functions, thereby inducing intrinsic positive correlations between cognitive functions. One step further back, each gene expressed in the brain (e.g., genes that code for neurotrophic factors, transcription factors, and any molecule involved in neurotransmission) typically influences several brain functions and properties, thereby inducing positive correlations between them. In parallel, many environmental factors (e.g., nutrition, socioeconomic status, education, diseases, and so on) influence more than one brain or cognitive function, thereby inducing further correlations. Finally, van der Maas et al. (2006) have shown that positive correlations between cognitive functions may emerge through mutual interactions in the course of cognitive development, even in the absence of intrinsic correlations. Thus, all of the factors underlying test performance are pleiotropic and conspire to produce positive correlations at all levels of description, hence the emergence of the positive manifold.

Note that, according to the explanation given previously, the positive manifold can arise in an entirely modular mind (because modules selected for different purposes nevertheless have to share underlying factors), and therefore there is no antagonism between modularity and general intelligence. Furthermore, the very same pleiotropic mechanisms are at work in other species and, therefore, readily explain that a *g* factor can be measured in nonhuman primates, rodents, and probably all organisms with a nervous system. Finally, in the speciation process, genes that progressively diverge between two populations influence more than one brain and cognitive function; therefore, the two populations are bound to eventually differ in more than one brain and cognitive function. This directly predicts that performance in different tests should covary across species, or what the authors term *G*. Thus, all of the evidence that the authors gather in support of a reified notion of general intelligence is more parsimoniously explained by the pleiotropy of the underlying factors, within and across species. The "independent evolution of large numbers of modules instead of general intelligence" is not "particularly difficult to reconcile with interspecific findings of *G*" (sect. 2.5, para. 5); it directly follows from an understanding of what modules are made of: the same building blocks, shared between species.

There is, therefore, no need to postulate that the positive manifold reflects one particular cognitive function or one brain function, whose evolution would require a special explanation. The positive manifold emerges spontaneously from the pleiotropy of all of the underlying factors. Only these underlying factors require an evolutionary explanation. It is indeed very interesting to inquire about the evolution of genes involved in brain development and function, the evolution of brain functions and properties, and the evolution of cognitive functions. If there is any brain or cognitive function whose evolution is a major puzzle, then it should be identified and studied as such. However, this is not the case for general intelligence, which does not reduce to a single brain or cognitive function, and whose evolution follows directly from that of the underlying biological, cognitive, and environmental factors.

General intelligence does not help us understand cognitive evolution

doi:10.1017/S0140525X16001771, e218

David M. Shuker,[a] Louise Barrett,[b] Thomas E. Dickins,[c] Thom C. Scott-Phillips,[d] and Robert A. Barton[d]

[a]*School of Biology, University of St Andrews, St Andrews KY16 9TH, United Kingdom;* [b]*Department of Psychology, University of Lethbridge, Lethbridge, Alberta T1K 3M4, Canada;* [c]*School of Science & Technology, University of Middlesex, London NW4 4BT, United Kingdom;* [d]*Evolutionary Anthropology Research Group, Durham University, Durham DH1 3LE, United Kingdom.*

dms14@st-andrews.ac.uk https://insects.st-andrews.ac.uk/
louise.barrett@uleth.ca http://directory.uleth.ca/users/louise.barrett
T.Dickins@mdx.ac.uk https://www.mdx.ac.uk/about-us/our-people/staff-directory/profile/dickins-tom
t.c.scott-phillips@durham.ac.uk https://thomscottphillips.wordpress.com
r.a.barton@durham.ac.uk
https://www.dur.ac.uk/anthropology/staff/academic/?id=122

Abstract: Burkart et al. conflate the domain-specificity of cognitive processes with the statistical pattern of variance in behavioural measures that partly reflect those processes. General intelligence is a statistical abstraction, not a cognitive trait, and we argue that the former does not warrant inferences about the nature or evolution of the latter.

Is "the presence of general intelligence" the "major evolutionary puzzle" that Burkart et al. claim? Like much of the literature on general intelligence in animals, the target article draws inferences about the nature and evolution of cognitive traits from the correlations among measures of performance, both within and between species. The "positive manifold" (sect. 1.1, para. 1) is thus taken to be a nontrivial finding, and *g* is treated as being – or reflecting – a trait with causal effects (a mechanism). *g*, however, is of course a statistical construct: When the authors refer to "the structure of cognition" (sect. 1.1.1, para. 1), what they actually describe is the statistical structure of variance in performance on behavioural tests. What can this statistical structure tell us about cognitive traits? We suggest that it tells us very little, or possibly nothing, because of the multiple plausible ways in which it might arise. Moreover, the analysis of *g* fails to provide a clear framework for empirical research, because the putative underlying mechanism, general intelligence, cannot be meaningfully defined in the absence of the correlations that are used as evidence for its existence.

More specifically, the reification of *g* involves a conflation of the proposed domain-generality of cognitive processes with the statistical pattern of variance in the behavioural output of those processes. Thus, "Massive modularity would appear to be irreconcilable with general intelligence" (sect. 1.2.1, para. 4) – well, only in the sense that apples are irreconcilable with oranges. Burkart et al. follow many in assuming that the positive manifold can be explained "by positing a dominant latent variable, the *g* factor, associated with a single cognitive or biological process or capacity" (van der Maas et al. 2006, p. 842). As pointed out by the latter authors, other explanations, which account for not only the presence of *g* but also its heritability and neuro-anatomical correlates, are not only possible, but also plausible. In citing van der Maas et al. (2006), Burkart et al. explicitly "equate general intelligence with the positive manifold" (sect. 1.1.1, para. 3), implying that their position and that of van der Maas et al. are in harmony. The point emphasised by van der Maas et al., however, and the point we also emphasise, is that the positive manifold provides little or no constraint on the possible architectures of cognition.

To labour the point, correlated variance does not imply any particular kind of cognitive process. That said, we might still want an explanation for why performance or behaviours are correlated across domains. Here, in brief, are some possibilities.

(1) They are not really different domains. For example, Reader et al. (2011) and Fernandes et al. (2014) found positive correlations among the rates of social deception, social learning, innovation,

extractive foraging, dietary breadth, percentage of fruit in the diet, and tool use across primate species, leading both sets of authors to conclusions about the domain-generality of cognitive processes. Neither these authors nor Burkart et al. explain how a domain is to be identified, and therefore how these behavioural measures can, in principle, be used to test for domain-generality. We can envisage plausible arguments to the effect that at least some of these behaviours draw on the same domain-specific processes. It is a question of natural ontologies: How do we carve nature at her joints? The only way that makes sense to us is in an evolutionary context where we identify a domain with a selection pressure. Deciding that "social" and "non-social" are distinct domains is, therefore, a hypothesis about what selection pressures have operated, not necessarily a fact about the world. Burkart et al. recognise this problem ("The issue of task selection is thus closely linked to the identification of domains in animal cognition" [sect. 2.4.2, para. 5]) but do not offer a convincing solution.

(2) Related to (1), it may well be that the behaviours measured are influenced by cognitive processes they share in common, but this does not mean it is helpful to describe those processes as "general processes," or that together they comprise "general intelligence." For example, primate species vary in their sensory-motor adaptations – in particular, in their stereo visual acuity and manual manipulative abilities – and these differences correlate with the evolution of binocular convergence supporting stereo vision, the size of visuomotor structures in the brain, and consequently overall brain size (Barton 2012; Heldstab et al. 2016). Clearly, such sensory-motor specializations may influence performance of a range of behaviours and/or experimental test procedures. Yet, describing them as "domain general" tells us nothing about how they work or how they evolved. We also do not share the optimism of Burkart et al. that reversal learning is free of such problems.

(3) Niche dimensions tend to be correlated (Clutton-Brock & Harvey 1977). For example, folivorous primates generally live in smaller social groups, have smaller home ranges, and engage less in extractive foraging and tool use than more omnivorous primates. Cognitive adaptations for specific niche dimensions could therefore theoretically be completely informationally encapsulated and yet performance across domains would still be correlated.

(4) The rates of naturally occurring behaviours in the wild (Reader et al. 2011; Fernandes et al. 2014, cited by Burkart et al.), may be systematically biased, leading to spurious correlations. Although these studies attempt to control for observation effort, they don't control for the number of individuals under observation. Rates of all behaviours will, other things being equal, correlate positively with group size and therefore with each other, because more individuals are under observation per unit time in larger groups. Variation in observability due to habitat will only exacerbate the problem. The implications are obvious.

For a theory to be useful, it has to be well defined in such a way as to generate testable predictions that differentiate it from other theories. Burkart et al., along with the wider literature on general intelligence and *g*, fail to achieve this. If we are to make progress in our efforts to understand the evolution and structure of cognition, we need to stop confusing the map for the territory.

It's time to move beyond the "Great Chain of Being"

doi:10.1017/S0140525X16001783, e219

Robert J. Sternberg

Department of Human Development, Cornell University, Ithaca, NY 14853.

robert.sternberg@gmail.com www.robertjsternberg.com
robert.sternberg@cornell.edu

Abstract: The target article provides an anthropocentric model of understanding intelligence in nonhuman animals. Such an idea dates back to Plato and, more recently, Lovejoy: On Earth, humans are at the top and other animals at successively lower levels. We then evaluate these other animals by our anthropocentric folk theories of their intelligence rather than by their own adaptive requirements.

Burkart et al. have written a very interesting, erudite, and anthropocentric account in the target article of how principles discovered for human intelligence might be generalized to animals other than humans. The presuppositions behind this article are captured well by Lovejoy (1936) in his book, *The Great Chain of Being*. The general idea, which goes back to Plato and Aristotle, is that there is a Great Chain of Being containing, among other entities, God at the top, then humankind, and then successively lower animals. At the top of the Earthly beings are humans. So if we want to understand other organisms, according to this view, we can do so by comparing them to humans and seeing in what ways they are similar and in what ways they are different and lacking. Much of early comparative psychology was based on this idea (e.g., Bitterman 1960).

Other areas of psychology and other behavioral sciences have not been immune from the logic of the Great Chain of Being, except that, in some cases, they viewed different cultures or races of people as occupying differentially elevated positions on the Great Chain (Sternberg 2004; Sternberg et al. 2005). Many eminent behavioral scientists, such as Sir Francis Galton and Raymond Cattell, believed in some version of the Great Chain (see https://www.splcenter.org/fighting-hate/extremist-files/individual/raymond-cattell). Moreover, traditional cross-cultural psychological studies of intelligence involved (and still involve) translating Euro-centric intelligence tests, such as the Wechsler, and then administering them to people in other cultures (e.g., Georgas et al. 2003).

But in the field of cultural studies of intelligence, progress has been made, largely due to the pioneering work of Luria (1976). Luria, in testing individuals in non-European cultures, found that the problems that were alleged to measure intelligence in European populations did not do so in other cultures because the individuals did not accept the presuppositions of the problems they were given. For example, when Uzbekistan peasants were given a syllogisms problem, such as, "There are no camels in Germany. The city of B. is in Germany. Are there camels there or not?", subjects could repeat the problem precisely and then answer "I don't know. I've never seen German villages …" The subjects did not accept the problems in the abstract modality for which they were intended. Of course, one could argue that they could not do so. But then, Cole et al. (1971) found that Kpelle tribesmen seemed not to be able to sort items categorically but rather sorted only functionally, until they were told to sort the way a stupid person would, at which point they had no trouble sorting categorically. In our own research (see Sternberg 2004), we found that rural Kenyan children and rural Alaskan Yup'ik Eskimo children could do tasks that were extremely important for adaptation and even survival in their own cultures (e.g., treating malaria with natural herbal medicines, finding their way across the frozen tundra from one village to another with no obvious landmarks) that their White teachers never could do, but were considered stupid by their teachers because they underperformed in school and on standard Euro-centric cognitive tests. Who was lacking intelligence: the children or the psychologists who gave them tests inappropriate to the demands of their everyday adaptation?

The tests we used for the Kenyan and Alaskan children cut to the heart of what intelligence is – ability to adapt to the environment. That is the core of intelligence, according to surveys of experts in the field of intelligence ("Intelligence and Its Measurement" 1921; Sternberg & Detterman 1986). But the tests that Burkart and her colleagues have devised are not tests highly relevant to animal adaptation; at best, and even then questionably, they are tests of folk conceptions of what animal intelligence should be from a human viewpoint.

An appropriate way to look broadly at the intelligence of any organism is to look at how well it adapts to the range of

environments it confronts. Gibson's (1979) concept of an affordance – an action possibility latent in the environment – is perhaps key here. The humans and other animals that are intelligent, in this view, are those that adapt well to the challenges of the range of environments they can encounter over the course of their lives. To understand animal intelligence, we should be looking at skills that are relevant to the animals' everyday adaptation, such as how well they can forage for food, create adequate shelter, and most important, avoid predators, including humans and the traps humans set for them.

Perhaps, furthermore, we humans should test human intelligence not with the often trivial tests we use (Sternberg 1990), but rather with tests of how well we humans can avoid the traps – for example, global warming, violence, pollution, poverty, inequality – that we set for ourselves.

Humans, with the serious problems they have created for themselves – pollution, global warming, weapons of mass destruction, terrorism, inequality, among others – may not be well positioned to be the judges of what intelligence looks like in other organisms, or of how intelligent they are. To hold other various animals to the standards of human folk conceptions of intelligence is perhaps an act of intellectual hubris. In the end, how intelligent, really, is a species that may be the only species ever to live on Earth actually to create and sow the seeds for its own destruction (Sternberg 2002)? If nonhuman animals were to create tests of intelligence for humans, perhaps they would create tests that would measure which humans were not intent on destroying both the animals' habitats and their own.

Disentangling *learning* from *knowing*: Does associative learning ability underlie performances on cognitive test batteries?

doi:10.1017/S0140525X16001795, e220

Jayden O. van Horik and Stephen E. G. Lea

Centre for Research in Animal Behaviour, Washington Singer Laboratories, Department of Psychology, College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4QG, United Kingdom.

jayden.van.horik@gmail.com www.jayden-van-horik.blogspot.co.uk
s.e.g.lea@exeter.ac.uk http://people.exeter.ac.uk/SEGLea/

Abstract: Are the mechanisms underlying variations in the performance of animals on cognitive test batteries analogous to those of humans? Differences might result from procedural inconsistencies in test battery design, but also from differences in how animals and humans solve cognitive problems. We suggest differentiating associative-based (*learning*) from rule-based (*knowing*) tasks to further our understanding of cognitive evolution across species.

In the target article, Burkart et al. highlight the importance of identifying variations in domain-general intelligence across species. However, with the exception of mice and possibly primates, there remains little evidence suggesting that variations in domain-general intelligence (*g*) underlie intraspecific variations in cognitive performance in nonhuman animals. Moreover, such an attribution remains debatable as procedural differences in test battery design may confound interpretations of the underlying mechanism. Our concern is exacerbated where support for a *g* factor is sparse and limited to studies that predominantly test subjects in the wild. For example, the mechanisms underlying success on test batteries designed to assess performances of birds in the wild (Isden et al. 2013; Keagy et al. 2011; Shaw et al. 2015) bear little resemblance to those effective in tasks presented to non-avian species tested in captivity (Herrmann et al. 2010b).

To accurately address whether it is meaningful to talk about domain-general intelligence in animals, it is important that the inherent design of the items within a cognitive test battery accurately capture domain-specific cognitive abilities, independent of procedural factors, and that relevant testing paradigms are used to assess the cognitive performances of subjects in the wild as well as in captivity. Direct comparisons between species are unavoidably difficult as different animals possess different adaptive specialisations; for example, a human cognitive test battery may assess verbal skills whereas nonhuman test batteries cannot. Test batteries, therefore, also need to consider the inherent differences in cognitive processes between species.

Performances on nonhuman cognitive test batteries, particularly those presented to subjects in the wild, require individuals to first interact with a novel apparatus before experiencing its affordances. Accordingly, such test batteries often use tasks that involve trial-and-error learning to quantify subjects' performances and assess their ability to learn to attend to cues based on reward contingencies. For example, subjects may be presented with tasks that assess how quickly they can learn to differentiate rewarded from unrewarded colours, or learn about the spatial location of concealed rewards. Although performances on such tasks are considered to capture domain-specific abilities, success will inevitably also be mediated by fundamental processes of learning that are common to the inherent design of these problems. As a result, an individual may perform well when learning both colour and spatial discrimination problems, not because this individual excels in anything we would want to call intelligence but because it is a relatively rapid learner of all kinds of association, including those involved in the two novel problems. Hence, what seems to be evidence for domain-general intelligence may reflect individual consistency in speed of associative learning, rather than individual consistency in cognition across different domains.

Between-species comparisons may be further confounded because associative learning ability plays a greater role in task performance in animals than it does in humans, and may play a greater role in some nonhuman species than others. Such differences may be particularly pronounced between evolutionarily disparate species such as primates and birds. Pigeons consistently show purely associative solutions to problems that humans, and to some extent nonhuman primates, tend to solve by the use of rules (e.g., Lea & Wills 2008; Lea et al. 2009; Maes et al. 2015; Meier et al. 2016; Smith et al. 2011; 2012; Wills et al. 2009). In humans, preferential attention to rules may expedite performances on rule-based tasks (Danforth et al. 1990), but may also impair responses to experienced contingencies (Fingerman & Levine 1974; Hayes et al. 1986). Consequently, as different cognitive processes govern the performances of different species on psychometric test batteries, analogous performances between human and nonhuman animals may be difficult to capture.

To overcome these issues, we highlight the importance of differentiating between performances on tasks that require subjects to *"learn"* to solve a problem, from performances on tasks that assess whether subjects *"know"* the solution to a problem. We therefore advocate the use not only of associative tasks such as discrimination learning of colour cues that require trial-and-error experience to solve, but also of tasks that require subjects to be trained beforehand to a particular learning criterion, so that their performance on a subsequent novel test or "generalization" condition can be assessed. Such conditions provide a controlled version of the tests of "insightful" or "spontaneous" problem solving that, from the time of Köhler (1925) on, have often been considered critical in assessing animal intelligence.

Learning tasks are particularly relevant when assessing individual differences in associative performances and may be more relevant when investigating the cognitive performances of nonhuman animals. Binary discriminations involving spatial or colour cues can be presented to subjects and their rates of learning quantified across these different cognitive domains. Although rates of associative learning may differ across domains (Seligman 1970), individual differences in such tasks may still be correlated, leading to a general factor reflecting associative learning ability

(hereafter "*a*"). However, for reliable comparisons, it also remains important to show that subjects' performances are consistent within domains.

Knowing tasks, by contrast, can be designed to assess the more flexible cognitive processes associated with rule-based learning or generalisation and may be more relevant when assessing cognition in humans. Such tasks require training subjects to a predetermined criterion of success to standardise their understanding of the problem, and then presenting subjects with a single test trial using novel cues. Importantly, performances on *knowing tasks* may highlight whether the mechanism underlying *g* in humans resembles that which may be found in nonhuman animals.

By incorporating both *learning* and *knowing* tasks into cognitive test batteries, we can address whether a general factor of cognitive performance in human and nonhuman animals is better represented by *g* or *a*. Distinguishing *learning* and *knowing* problems, therefore, provides a measure of individual variation in both domain-specific and domain-general abilities that do not just reflect speed of associative learning, and so can be used to assess whether variation in nonhuman cognitive performance reflects a dimension of general intelligence of the same kind as is thought to underlie human variation.

ACKNOWLEDGMENT

JvH was funded by an ERC Consolidator grant awarded to Joah Madden (616474).

A pointer's hypothesis of general intelligence evolved from domain-specific demands

doi:10.1017/S0140525X16001801, e221

X. T. (Xiao-Tian) Wang

Psychology Department, University of South Dakota, Vermillion, SD 57069.
xtwang@usd.edu http://usd-apps.usd.edu/xtwanglab/

Abstract: A higher-order function may evolve phylogenetically if it is demanded by multiple domain-specific modules. Task-specificity to solve a unique adaptive problem (e.g., foraging or mating) should be distinguished from function-specificity to deal with a common computational demand (e.g., numeracy, verbal communication) required by many tasks. A localized brain function is likely a result of such common computational demand.

The authors of the target article provide an excellent discussion on the evolution of human intelligence, particularly on the formation of secondary modules that are more variable and domain-general. As discussed in section 1 of the target article, general intelligence seems evolutionarily implausible because the mind is populated by a large number of adaptive specializations that are functionally organized to solve evolutionarily typical and recurrent problems of survival and reproduction (see also Cosmides et al. 2010; Wang 1996). To resolve this paradox, the authors propose a model that construes the mind as a mix of truly modular skills (primary modules) and more variable and flexible skills (secondary modules) that are ontogenetically acquired via the guidance of general intelligence (see sect. 1.2.3). In the following, I propose a novel hypothesis to extend this discussion by showing that secondary or higher-order modules can be formed not only ontogenetically, but also phylogenetically as adaptations, evolved from domain-specific modules.

If general intelligence consists of a set of secondary modules, each secondary module may be an evolved programing solution for a function that could be shared by multiple primary modules. These secondary modules of general intelligence can be either ontogenetically constructed or phylogenetically evolved. Imagine that a computer architect was creating a system called *Unix* using the programming language C. At the beginning, the operating system was written in assembly, where nearly every line would contain memory addresses. Would it be possible to program the system for its input/output devices without repetitively stating these tedious memory addresses? This problem has been solved by creating a pointer variable, whose value specifies the address of a memory location. If a memory address is called upon repeatedly, creating a pointer to store the address would be an effective programing solution. Similarly, if a random number generator is used repeatedly by many local modules, it would be more efficient to make it globally accessible by each of the modules.

Now imagine you are using a computer and have created many folders for different papers. At the beginning, you included a copy of a word processor in each folder. You then realized that all of these papers require a word processor. It would be more efficient if you place a single copy of a generic word processor in a visible place that is accessible by all of the papers. This word processor has then become a general tool for a common requirement of different tasks. Similarly, numeracy, as a component of general intelligence, may be evolved as a result of a common demand by multiple specific adaptations (e.g., counting foraging outcomes; gauging social exchanges, assessing mate values, tracking reciprocal activities, etc.). A general-purpose device would be cognitively economical if it is utilized for multiple tasks. From a design viewpoint, general intelligence comes as a solution for overlapping components of primary modules or for coordinating secondary modules via executive functions (see sect. 1.2.2). From this perspective, task-specificity to solve a unique adaptive problem (e.g., foraging, hunting, or mating) should be distinguished from function-specificity to deal with a common computational demand (e.g., numeracy, verbal communication, etc.)

By the same token, if a particular emotion is a common component of many specific adaptations, this basic emotion would become a general mechanism shared by these adaptations. For instance, anger is the expression of a neurocomputational system that evolved to adaptively regulate behavior in the context of resolving conflicts of interest in favor of the angry individual (Cosmides & Tooby 2013). Anger can be triggered by multiple task-specific adaptations, such as territory defense, mating competition, sibling rivalry, and cheater detection. Once triggered, the anger system would produce one of two outputs: threatening to inflict costs (aggression) or threatening to withdraw expected benefits (Cosmides & Tooby 2013). Similarly, fear is a basic emotion that plays a role in multiple adaptations and has its brain center mainly located in the amygdala. This localized brain function allows the organism to react not only to specific and typical fear-inducing stimuli, but also to learn to react to nonspecific stimuli with fear via fear conditioning (e.g., Phelps & LeDoux 2005).

General intelligence and basic emotions may both be solutions for multiple primary modules that demand some common functions. This pointer's hypothesis of general intelligence challenges a couple assumptions in the research literature of cognitive evolution. As indicated by the authors, many previous accounts of evolution of human intelligence assume that domain-specific modules ought to be cheaper and simpler than domain-general cognitive mechanisms (see sect. 1 for relevant discussion). However, being specific does not necessarily mean that the mechanism is simple or cognitively economical. Because a domain-specific mechanism is designed for solving a specific problem, its design purpose is to do whatever it takes to solve the problem instead of achieving structural simplicity, computational economy, or functional efficiency. Such designs can be either as exquisite as the human visual system or as patchy and lousy as a male's reproductive system, revised and modified from the Wolffian duct. Thus, these adaptive specializations can either be cheap and simple or costly and complex. Unlike engineering designs, evolutionary designs cannot afford to erase existing blueprints and start from scratch. Evolutionary efficiency is inevitably an efficiency under phylogenetic constraints.

The pointer's hypothesis also challenges the notion that domain-specific mechanisms are more localized in the brain than domain-general mechanisms (see also sect. 1.1 for relevant discussion). However, this notion is at odds with the following two observations. First, a specific adaptation can be implemented by a distributed neural network. Second, a localized brain function is likely a result of a common demand of multiple primary modules. Thus, a more general-purpose mechanism may be implemented by allocating a particular brain region to perform a function shared by multiple primary modules. For instance, a localized motor cortex (e.g., the precentral gyrus) can be used for motor controls in foraging, hunting, gathering, mating competition, and so on. For the same reason, localized brain regions for language processing serve as a general-purpose system for all of the tasks that require information exchange and verbal communication.

When does cultural transmission favour or instead substitute for general intelligence?

doi:10.1017/S0140525X16001813, e222

Andrew Whiten

University of St Andrews, Centre for Social Learning and Cognitive Evolution, School of Psychology and Neuroscience, University of St Andrews, St Andrews KY16 9JP, United Kingdom.

aw2@st-andrews.ac.uk **www.st-andrews.ac.uk/profile/aw2**

Abstract: The cultural intelligence hypothesis is an exciting new development. The hypothesis that it encourages general intelligence is intriguing, but it presents a paradox insofar as social learning is often suggested to instead reduce reliance on individual cognition and exploration. There is thus a need to specify more clearly the contexts in which cultural transmission may select for general intelligence.

Burkart et al. provide a comprehensive and erudite review offering new perspectives on the burgeoning developments in human and nonhuman animal research on intelligence. I particularly welcome the eventual focus on the relatively recently formulated cultural intelligence hypotheses, which I have collaborated in developing (Whiten & van Schaik 2007; and see Whiten [in press] for relationships with the broader, earlier research on social intelligence). However, I have a few comments and puzzlements to share.

The first concerns the authors' conclusion in section 1.2.2 that "natural selection for social learning seems to automatically trigger selection on individual learning and general cognitive ability, suggesting that ontogenetic canalization through social learning may have contributed to enabling the evolution of domain-general cognition" (para. 8). This principle is elaborated further in section 3.3, hypothesising that selection for cultural intelligence offers an explanation for the evolution of greater general intelligence in some species, largely because the rewards consequent on the efficiency of learning from experienced others minimise energetic constraints on encephalization required for greater general intelligence. However, this is an "enabling" explanation rather than one positing positive selection on general intelligence through an emphasis on cultural transmission, and I suggest there is something of a paradox here, or at least a conundrum.

The conundrum is that Burkart et al. propose that cultural learning *encourages* general intelligence, whereas it is common in the social learning literature to assert, to the contrary, that a core adaptive advantage of this form of learning is that it *reduces* the costly needs of individual learning. Thus, for example, it appears from the restriction of chimpanzees' nut-cracking to only far West Africa that most chimpanzees have insufficient general intelligence to invent the practice, despite availability of the requisite raw materials (excepting at least one rare innovator, at some stage); however, a suite of experiments has shown that naïve chimpanzees (some from East Africa) can learn the skill following observation of a nut-cracker (Whiten 2015). This suggests that most wild chimpanzees in the West achieve the skill via observational learning, removing selection pressure on the general intelligence necessary to invent the skill.

If this is the case, it suggests that Burkart et al. have more work to do to specify just exactly what aspects of general intelligence they propose may be selected for in such scenarios. They mention practice in this context, which is certainly protracted in the example of nut-cracking (Whiten 2015). But the practice involved in perfecting nut-cracking learned from others seems rather far from the definition that "general intelligence, as defined in either humans or nonhuman animals, stresses reasoning ability and behavioural flexibility" (sect. 1.1, para. 1). In the human case, the phenomenon of "over-imitation," in which children (apparently unlike other apes) copy others' visibly causally irrelevant actions suggests a marked relinquishing of reasoning and flexibility, commonly interpreted as a correlate of our species' extreme reliance on cultural transmission (Whiten et al. 2009).

Does the authors' emphasis on the potential knock-on effects of cultural intelligence on general intelligence perhaps neglect the direct effects of selection for cultural transmission encouraging other, socio-cognitive enhancements with implications for encephalization? The cultural intelligence hypothesis was originally developed to explain the encephalization and intelligence of the great apes (Whiten & van Schaik 2007), which was not accounted for by broader social intelligence theories that work well for primates in general (Dunbar & Shultz 2007a). Consistent with this, a recent study reported multiple-tradition cultures for gorillas (Robbins et al. 2016) that are consistent with those earlier described for chimpanzees (Whiten et al. 1999) and orangutans (van Schaik et al. 2003) and appear rich compared to the putative cultures of other animals, although a parallel analysis for spider monkeys, in some ways a New World chimpanzee-like niche, reported a quite similar complexity (Santorelli et al. 2011), and studies of capuchin monkeys suggest something similar may await systematic assessment (e.g., Coehlo et al. 2015) . Both the latter species are relatively encephalized, and of course the same is true for cetaceans for which a strong evidential case has been made for multiple-tradition cultures including foraging techniques, migration routes, and song (Whitehead & Rendell 2015). The social learning capacities of such animals with heavy dependence on extended cultural repertoires may themselves need to be cognitively sophisticated, including imitative and emulative processes, with neural demands (Whiten 2017; in press). In addition, encephalization may be extended simply to facilitate the storage of a greater cultural repertoire. In the human case, this may be very significant when one contemplates the vast scope of the cultural information we assimilate, from language to all aspects of social and material culture.

The latter leads to a related but different comment. The authors tend to run together two threads in the literature when referring to "the cultural intelligence hypothesis" (sect. 3.3, para. 3), and I think it would reduce potential confusion to separate these. The first thread is exemplified by the writings of Tomasello et al. that are cited, such as Moll and Tomasello (2007). The second thread is exemplified by the writings of van Schaik et al. (e.g., Whiten & van Schaik 2007; van Schaik & Burkart 2011). It is this second thread that sets out a cultural intelligence hypothesis addressed originally to the problem of great ape intelligence and encephalization, but in principle relevant to any relevant animal species. By contrast, the first thread was specifically concerned with what makes humans different from all other animals, and was originally and appropriately dubbed "the Vygotskian intelligence hypothesis" (Moll & Tomasello 2007, p. 639). This made sense to me, until Herrmann et al. (2007) then referred to these ideas as "the cultural intelligence hypothesis" (p. 1360). This was potentially quite confusing insofar as the argument was that it did not apply to nonhuman species. I feel it is important to recognise these differences, whether that is achieved by reverting to the "Vygotskian" tag to distinguish the "human" focused

version. This is not, of course, to deny that there is a potentially important linkage between the sets of ideas embedded in these two theories.

General intelligence is a source of individual differences *between* species: Solving an anomaly

doi:10.1017/S0140525X16001825, e223

Michael A. Woodley of Menie,[a,b,1] Heitor B. F. Fernandes,[c,1] Jan te Nijenhuis,[d] Mateo Peñaherrera-Aguirre,[e] and Aurelio José Figueredo[c]

[a]*Scientist in Residence, Technische Universität Chemnitz, 09111 Chemnitz, Germany;* [b]*Center Leo Apostel for Interdisciplinary Studies, Vrije Universiteit Brussel, Brussels 1050, Belgium;* [c]*Department of Psychology, University of Arizona, Tucson, AZ 85721;* [d]*Department of Work and Organizational Psychology, University of Amsterdam, 1000 GG Amsterdam, The Netherlands;* [e]*Department of Psychology, University of New Brunswick, Fredericton, NB E3B 5A3, Canada.*

Michael.Woodley@vub.ac.be hbffernandes@gmail.com
nijen631@planet.nl Mateo.PA@unb.ca ajf@u.arizona.edu

Abstract: Burkart et al. present a paradox – general factors of intelligence exist among individual differences (*g*) in performance in several species, and also at the aggregate level (*G*); however, there is ambiguous evidence for the existence of *g* when analyzing data using a mixed approach, that is, when comparing individuals of different species using the same cognitive ability battery. Here, we present an empirical solution to this paradox.

As Burkart et al. note in the target article, it is possible that the *g* factors that exist within species at the level of individual differences have somewhat different factorial characteristics for each species. For instance, certain cognitive elements that combine to give rise to *g* in chimpanzees may fall outside of the positive manifold in other species (e.g., humans). In other words, performance in certain abilities may be driven by *g* in some species but not in others. Lack of *measurement invariance* (i.e., discordance between species in terms of which cognitive abilities give rise to *g*) renders single batteries unable to identify a *g* factor common to individuals of different species (i.e., the mixed approach). One cause of these compositional differences may be the different ways in which ancestral selection pressures shaped the *g* factors across different species. Some species may have highly integrated abilities, dominated by a strong *g* factor, whereas others might have highly specialized and largely independent abilities, where the positive manifold of correlations underlying *g* is weaker.

Another potentially significant cause of the failure of measurement invariance across individuals of different species may be *floor* or *ceiling* effects upon performance. For example, a cognitive task that may be hard for one species may be trivially easy for another, more intelligent species. The latter condition is characterized by all or most individuals performing maximally well, revealing a ceiling effect. Hence, the *g* loading of the success rate at solving this task may be high for the less intelligent species, but will be low for the more intelligent one – this species having hit the test ceiling.

Operationally, both (1) species-specific specialization or modularization of cognitive abilities and (2) floor/ceiling effects can be identified empirically based on within-species statistical distributions in performance. The two conditions are likely to share a common observable feature: that is, low within-species variability in certain tasks. Highly specialized abilities are proposed to be species-typical and monomorphic, with little to no interindividual variation (Tooby & Cosmides 1990). Consistent with this, human and nonhuman primate data indicate that cognitive functions that are more specialized (and thus less *g*-loaded) exhibit lower phenotypic and genetic variability (Spitz 1988; Woodley of Menie et al. 2015). The presence of ceiling or floor effects in measurement when testing abilities in a given species also, by definition, limits variation. These alternative scenarios are therefore connected, as any apparent floor or ceiling effect in the performance of modularized abilities may not be due to a poor measurement approach but, rather, due to adaptive species-typical modularization.

We propose that the mixed design would support the presence of a *g* factor inclusive of individuals of different species if species differences in cognitive ability are larger on tasks that share more variance with others (larger part-whole correlations, representing *g*-loadings) but not if species differences are uniform across tasks. Here, we use combined data from two sources (Herrmann et al.

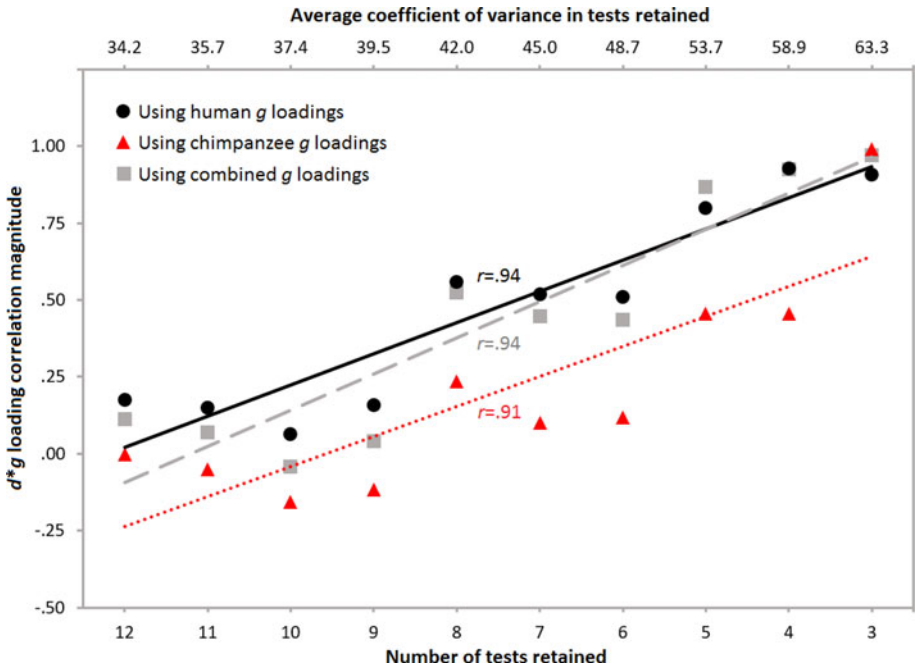


Figure 1 (Woodley of Menie et al.). Increasing magnitude of the vector correlations between task *g* loadings and the difference scores (*d*) between human children and chimpanzee performance, as a function of the average coefficient of variance of the tests kept in analyses.

2010b; Woodley of Menie et al. 2015) on the Primate Cognitive Test Battery (PCTB; Herrmann et al. 2007) performance in human children and chimpanzees to test this hypothesis and examine the importance of the confounding role of tasks on which individuals of at least one of the two species exhibit limited variability in scores.

Human children outperform chimpanzees to a greater degree on more *g*-loaded PCTB tasks – this can be demonstrated by correlating the vector of task *g* loadings with the vector of the between-species differences in performance (*d*) on those same tasks. To examine whether the true strength of this relationship was masked by the inclusion of tasks that yielded little within-species variation, we eliminated tasks from the analyses sequentially, starting with those that yielded the smallest coefficients of variance (CV) in human performance. The relationship between *g* loadings and the size of human-chimpanzee differences was thus examined in multiple stages, with each successive step having a more stringent cutoff for CV. Recall that the ceiling effects are a feature of the ease with which humans can execute certain basic cognitive tasks, suggesting that these abilities are modularized in human populations. CV was in fact smaller among humans on all tasks, suggesting that they solved all tasks more easily than chimpanzees.

Figure 1 shows that the g^*d vector correlation magnitude increased inversely to the number of tasks retained, with smaller numbers of tasks exhibiting larger variation among the human participants yielding bigger vector correlations. The vector correlation magnitude approached unity when only the three tests with the highest human CV values were used. The association was indifferent to the use of different *g* loadings (human, chimpanzee, and averaged) as the basis for computing the g^*d vector correlations (the correlations between the vector correlation magnitudes and average CV across tasks ranged from .91 to .94, $p < .05$).

Furthermore, as expected, tasks yielding smaller CV values were also less *g*-loaded in humans ($r = .52$; one-tailed $p < .05$), which replicates prior findings involving chimpanzees (Woodley of Menie et al. 2015).

This approach is currently being applied by our group to comparisons involving a larger number of species. The implication of our finding is that differences between individuals of different species may be consistently concentrated on *g* – this being especially apparent when focusing on experimental tasks whose design permits sufficient within-species variation. This finding furthermore indicates that the patterning of species differences in the *g* and *G* factors are concordant, meaning that they are likely one and the same, reinforcing the arguments put forward by Burkart et al.

NOTE

1. The two first authors contributed equally to this commentary.

Authors' Response

Future directions for studying the evolution of general intelligence

doi:10.1017/S0140525X17000024, e224

Judith M. Burkart, Michèle N. Schubiger, and Carel P. van Schaik

Department of Anthropology, University of Zurich, CH-8125 Zurich, Switzerland

Judith.Burkart@aim.uzh.ch

http://www.aim.uzh.ch/de/Members/seniorlecturers/judithburkart.html

michele.schubiger@aim.uzh.ch

http://www.aim.uzh.ch/de/Members/phdstudents/micheleschubiger.html

vschaik@aim.uzh.ch

http://www.aim.uzh.ch/de/Members/profofinstitute/vanschaik.html

Abstract: The goal of our target article was to lay out current evidence relevant to the question of whether general intelligence can be found in nonhuman animals in order to better understand its evolution in humans. The topic is a controversial one, as evident from the broad range of partly incompatible comments it has elicited. The main goal of our response is to translate these issues into testable empirical predictions, which together can provide the basis for a broad research agenda.

R1. Introduction

We are grateful to the commentators for raising a wide variety of issues. Because these generally fall into a number of distinct categories, we organized our response around them, as follows:

Section R2: Domains of cognition
Section R3: Tasks for test batteries
Section R4: Domain-specificity versus domain-generality
Section R5: What is *g*/*G*? (More) on the necessity of validation
Section R6: *g*/*G* and brain size
Section R7: *g* and biological fitness
Section R8: Cultural intelligence

Whenever commentators provided input in more than one of these categories, we discuss them in more than one of the sections. Some of the commentaries show that we did not always present our points with sufficient clarity, so we also take the opportunity to make these clarifications as well as to highlight what we did *not* claim.

The complicated nature of the issues is illustrated by the fact that different commentators made confident claims that are mutually incompatible. For instance, **Ramus** and **Arden & Zietsch** argue that the evolution of *g* is no puzzle at all, and that it is most likely present in all organisms with a nervous system, **Jacobs & Gärdenfors** similarly argue that it is hardly surprising or controversial that some individuals consistently perform well, and **Pepperberg** recites the impressive examples from bird cognition that to her necessarily imply the presence of general intelligence in at least some birds. On the other hand, other commentators question whether there is evidence for *g* or *G* in any nonhuman species at all (**Huber**) or feel the importance of *g* is overestimated (**Amici, Call, & Aureli [Amici et al.]**).

The goal of our target article was to make progress on understanding animal intelligence without getting bogged down in terminological debates on what exactly *g* represents. We suggested complementing the psychometric approaches, which are a necessary first step to establish the possibility of general intelligence, with a variety of validation measures and more demanding tests that look for domain-generality of cognitive processes. This is even more important because animal studies are unlikely to ever reach the sophistication in terms of tests and the sample sizes needed to attain the practical level of utility achieved by human intelligence testing (pace **Arden & Zietsch**). We will focus, therefore, especially on these forward-looking points.

R2. Domains of cognition

One of the unresolved issues in nonhuman psychometrics is what an ideal test battery should look like. The criteria are

obvious: (1) A test battery should be composed of tasks from a broad range of domains rather than tasks that resample performance in the same domain, and (2) the tasks should reliably assess the cognitive abilities they are supposed to assess, both within and across species.

To fulfill criterion (1), it is crucial to know what a domain is. In the classical psychometric tradition, a domain of mental ability refers to a statistically derived group factor on which a set of tasks show strong loadings. For instance, Deary et al. (2010) referred for humans to the group factors processing speed, memory, spatial ability, reasoning, and vocabulary. Thus, the straightforward approach to identify domains in nonhuman animals is to likewise use large batteries of tasks and identify the factor structure statistically. This bottom-up, data-driven approach in animals is ambitious because it requires very large sample sizes. However, this does not mean that, in the meantime, we are unable to correctly identify *g* in animals. Empirical tests with human subjects have revealed that as long as there is sufficient variety in the tasks of test batteries, *g* factors derived from different test batteries are almost perfectly correlated (Johnson et al. 2008).

Evolutionary approaches often consider domains to be functional contexts (see also **Wang**). These can be very specific such as finding food, finding mates, deterring predators, outwitting conspecifics, or cooperating with others (Figure R1), or rather broad, such as in the classic trichotomy in physical, spatial, and social cognition, as followed by Tomasello & Call (1997). Obviously, the individual may fail to show a particular ability that it is shown to possess in one context, when tested in a functional context different from the one for which it evolved. Thus, it remains challenging to identify domains.

Importantly, it is not useful to decide a priori that one notion of domain would be more correct than the other one. Rather, acknowledging that a domain can refer to both a mental ability (as in the psychometric tradition) and a functional context (as in evolutionary approaches; see also the proposal by **Shuker, Barrett, Dickins, Scott-Phillips, & Barton [Shuker et al.]** or **Hauser**'s examples of abilities that "*cut across domains*") allows us to ask the questions that are at the core for understanding the evolution of general intelligence: Under what circumstances can a species that evolves a cognitive ability for a specific context generalize this ability to other contexts as well (cf. Stevens et al. 2016)? Can this be true for some abilities, but not others? And are these the same species that also show correlated performance across mental abilities and thus show psychometric *g*?

Thus, rather than committing to one specific notion of a domain, it is better to identify ways to combine these perspectives to ultimately better understand the evolution of intelligence. Furthermore, this combined perspective will also clarify many of the issues surrounding domain-specificity and domain-generality, and modularity (sect. R4).

Due to the sample size problems in identifying domains using the psychometric approach, some have defined domains a priori (e.g., Herrmann et al. 2007). However, as we point out in the target article, empirical data do not necessarily confirm that tasks supposed to tap into one such domain actually also cluster around it (see sect. 2.4.1 in target article). Although we agree in principle with the proposal of **Amici et al.** that it is important to take multifactorial approaches into account (see also **Lee & Chabris**), we think more empirical work is needed to identify which tasks indeed measure particular cognitive operations (such as inference) or mental capacities (such as working memory). In fact, most tasks probably tap into more than one operation and/or capacity, and it therefore is highly unlikely that one specific test measures one specific cognitive ability, as also pointed out by **Huber** and **Ramus**. Even for humans there is often a lack of agreement on how to quantify specific cognitive constructs such as, for instance, working memory (Oberauer et al. 2005).

Finally, Figure R1 also helps us to answer how ecologically valid a test should be. An often-made claim is that in order to fully appreciate a given species' cognitive potential, one should look only at problems of high ecological relevance to this species (e.g., **Sternberg**). However, when

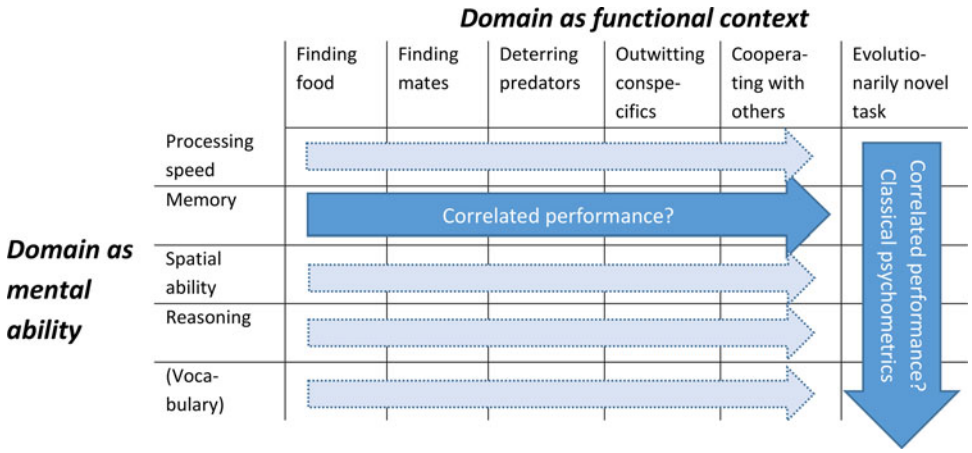


Figure R1. Domains can refer to statistically derived group factors such as processing speed, memory, spatial ability, reasoning, or vocabulary in humans (Deary et al., 2010) or to evolutionarily functional contexts. Classical psychometric studies (vertical arrow), in humans and nonhuman animals, typically refer to domain as mental ability. From an evolutionary perspective, however, it is equally informative to ask to what extent a given cognitive ability is correlated across functional contexts (horizontal arrows). Because performance across functional contexts is likely to vary, evolutionarily novel tasks are most suitable for classical psychometric studies that aim at testing individuals across mental ability domains.

we look only at ecologically relevant problems we will never be able to disentangle whether we are dealing with a primary module or with the result of true and flexible problem-solving ability. As indicated in Figure R1, there are at least two ways to demonstrate this kind of flexibility. A psychometric approach would compare different abilities in evolutionarily novel tasks, precisely in order to avoid tapping into predominantly hard-wired solutions to recurring adaptive problems. This approach is particularly powerful in combination with validation studies (sect. R5). The other one is to compare the same ability across functional domains (e.g., asking whether the excellent memory abilities in the ecologically relevant context of caching in food-caching birds also generalize to other functional contexts). Although this second possibility has received surprisingly little attention (but see Stevens et al. 2016), it presents an excellent opportunity to externally validate a psychometrically derived *g* factor (see also **Locurto,** sect. 2.5 in target article, and sect. R4).

R3. Tasks for test batteries

Many commentators have proposed specific tasks, and we think these are excellent proposals that will help construct broad and diverse test batteries. **Van Horik & Lea** stress the necessity to also include what they call *knowing tasks*, which can assess rule-based learning or generalization. Such knowing tasks, as for instance reversal learning, also have an additional advantage, linked to criterion 2 of a good test battery: that the tasks should reliably assess the cognitive abilities they are supposed to assess, both within and across species (sect. R2). The advantage is that knowing tasks are less vulnerable to producing variation in performance due to differences in sensory-motor specializations between species (see **Shuker et al**., and sect. 2.4.2 in the target article for a detailed discussion of Macphail's critique of species comparisons). We are confident that many of these issues can be resolved, particularly in closely related species, but also agree that it remains problematic for comparisons of very diverse species, such as cephalopods, with completely different sensory-motor traits and even body plans (**Holekamp & Miikkulainen**).

Commentators **Jacobs & Gärdenfors** highlight causal cognition tasks; **Huber**, reasoning by exclusion; and **Buskell & Halina**, tasks suggestive of de-coupled representations. Tasks in which birds excel may also be useful additions to existing test batteries. As stressed by **Pepperberg**, corvids and parrots are particularly prime candidates for exhibiting high levels of general intelligence, and we look forward to future empirical studies that explicitly address this question with psychometric test batteries combined with predictive validation studies (**Locurto**; see also sect. R5). **Pennycook & Rand** add the possibility to include studies of cognitive control and metacognition.

Commentators **Pennycook & Rand** also argue that variation in performance between individuals may not only reflect cognitive capacity per se, but also the willingness to engage in effortful cognitive processing when a simpler, more routine solution seems available. They thus refer to the concept of *need for cognition*, which is an individual predisposition rather than a cognitive ability, even though in humans it is correlated with general intelligence (Hill et al. 2016b). This link is entirely consistent with our view that cognitive skills emanating from general cognitive capacity (i.e., the downward pathway in Figure 3 of the target article) are ontogenetically constructed. Therefore, individuals with higher levels of need for cognition will more systematically expose themselves to situations that require effortful processing and are, therefore, more likely to establish a larger and more powerful set of cognitive skills. To what extent substantial variation in willingness to engage in effortful processing is also available in nonhuman animals remains to be established, but an increasing body of research suggests that this may account for the systematic differences in cognitive performance between captive and wild primates such as orangutans (van Schaik et al. 2016).

R4. Domain-specificity versus domain-generality

Several commentators (**Arden & Zietsch**; **Hauser**; **Jacobs & Gärdenfors**; **Martins & Di Paolo**) think we advocate a strict dichotomy between a mind made up of modules and one that has general intelligence. This is surprising, because in the target article, we devote an entire section (1.2.2 including Table 1) to exactly why such a dichotomy is not helpful for the debate, and in Figure 3 we present what we believe is the most promising working model given the current state of evidence on *g*/*G* in human and nonhuman animals. This working model suggests that the actual skill set of an individual can be construed as a mix of primary or secondary modules. Whereas both can be sensitive to experience (as detailed in Table 2 in the target article and unlike the claim by **Lewis, Al-Shawaf, & Anderson [Lewis et al.]**), this sensitivity differs: Primary modules are experience-expectant; secondary modules, experience-dependent.

The extent of domain-specificity and domain-generality obviously critically depends on the notion of domain one adopts (see sect. R2). This distinction between domain as mental ability or as functional context, as detailed in Figure R1, is also reflected in the many different notions of modularity that have been the focus in the modularity debate ever since Fodor. Proposals range from defining modules as adaptive behavioral functions localized in delimited neural areas to seeing them as being nothing else than components and processes into which a phenotype can be decomposed (Barrett 2015). When referring to domain specificity, we focus on functional specialization in the biological, ultimate sense (as used, for instance, in the commentary by **Amici et al.**), rather than referring to any other specific notion of domain-specificity (**Lewis et al.**). This allows us to ask under which conditions a cognitive adaptation to a specific problem, – for instance, the impressive memory of food-caching birds – can also be used in a different context. From comparative studies, we know that cognitive abilities in nonhuman animals sometimes do generalize to other domains, sometimes do so partially, and sometimes do not at all (Cauchoix & Chaine 2016; Stevens et al. 2016). When cognitive abilities do generalize, domain-general cognition emerges, and the fundamental question we are interested in is under what conditions this can happen.

If it is not straightforward to define domain-specificity or modularity, this is even truer for domain-generality in cognition or general intelligence. Not unexpectedly, several

authors have commented on this. **Sternberg** suggests that "an appropriate way to look at the intelligence of any organism is to look at how well it adapts to the range of environments it confronts" (see also **Bar-Hen-Schweiger, Schweiger, & Henik [Bar-Hen-Schweiger et al.]**). This is problematic because intelligence is only one way among many to adapt to the environment. Under this conception, the thickening of the fur in autumn clearly is an adaptation to the environment but most would agree it has nothing to do with intelligence. For the same reason, it is problematic to refer to domain-generality as phenotypic plasticity (**Buskell & Halina**) because it is merely one aspect of that. Equally unhelpful is the proposal by **Sternberg** to look at skills that are relevant to an animal's everyday adaptation. If we go back to our food-caching birds and assess only their memory capacities in the context of food caching, we are simply not able to exclude the possibility that we are dealing with a highly domain-specific capacity.

It is this latter issue that makes psychometric *g* studies so appealing to comparative psychologists. As highlighted in the target article (sect. 2.4 and sect. 2.5) and by several commentators, this approach comes with a whole suite of issues, addressed here in sections R5, R6, and R7. However, at least in humans, *g* is well established, known to reflect general intelligence as broadly defined in the target article (following Gottfredson 1997 and Nisbett et al. 2012), has a variety of genetic correlates, including those linked to neurobiological features, and is a good predictor for various measures of life outcome and thus shows high levels of predictive validity. In nonhumans, more validation is critically required, as also emphasized by **Locurto.**

R5. What is *g*/*G*? (More) on the necessity of validation

A common concern not only to many commentators, but also to ourselves (see sect. 2.5 in the target article), is that once one has statistically established a reliable *g* factor in a given species, based on adequate tasks from a broad range of domains, it does not necessarily follow that this corresponds to general intelligence as broadly defined and predictive for a variety of life outcome measures, as is the case in humans. Furthermore, even for humans, despite the predictive validity of *g*, it does not necessarily follow that the predominant hierarchical CHC model (see para. 1.1.1 in the target article) is the only compelling interpretation. Some suggest that a statistical *g* simply results from variation in non-cognitive factors, such as developmental stability, and thus reflects general health (**Lewis et al.**), or genetic load or heterozygosity (**Arslan, von Borell, Ostner, & Penke [Arslan et al.]**). Several commentators discuss plausible alternative models to the hierarchical CHC model. Thus, the process overlap theory of **Kovacs & Conway** (see also Kovacs & Conway [2016] and the responses in the same issue of *Psychological Inquiry*) argue that performance in a specific task is determined by several cognitive abilities, and different tasks resample the same abilities to different extents. The mutualism model of van der Maas et al. (2006) (see **Ramus**; **Shuker et al.**) argues that *g* arises because of mutual interactions during cognitive development. These alternatives thus see *g* as a composite of independent but overlapping or interacting processes, rather than as a single top-down ability (which **Bar-Hen-Schweiger et al.** see as the ability to engage in mental manipulation). Hence, they regard *g* more as a developmental consequence or emergent property, rather than as an underlying latent variable, and are fully compatible with the constructive nature of domain-general cognition.

Nonetheless, these varying interpretations still argue for the presence of some domain-general ability, even if it is entirely emergent, and remain compatible with the various neurobiological and genetic correlates of *g*. For instance, if *g* arises as a result of process overlap because cognitive tests from different domains tap the same domain-general executive processes such as working memory, this would exactly correspond to domain-generality as identified by the horizontal arrow in Figure R1.

The exact nature of *g* continues to be a major puzzle (Deary et al. 2016). But we would argue that this puzzle doesn't have to be fully solved for studies of domain-general processes to be useful, provided *g* in nonhuman animals survives future tests of external validation, as we discuss next. Some argue that emergent properties or statistical constructs cannot be meaningful individual-level variables subject to natural selection (**Arden & Zietsch**; **Ramus**). We do not think this is correct, because even if *g* is an emergent property of several cognitive and even non-cognitive processes, natural selection should evaluate the degree to which these processes are compatible or even how they synergize to produce organized and appropriate behavior. Indeed, each of the models for *g* discussed previously is compatible with natural selection having honed *g*. Hence, considering *g* a measurable property of an organism is not reification.

These ambiguities in the interpretation of *g* necessarily spill over into interpretations of animal studies, especially in light of the weaknesses of intraspecific studies, as highlighted by **Arden & Zietsch**, **Lee & Chabris**, and **Lewis et al**. Nonetheless, as we have argued in section 2.4 of the target article, the fact that all purely interspecific studies produce clear evidence for *G* is incompatible with the notion of *g* as an artifact of testing the same basic ability or of inadequate test batteries. One can, of course, criticize the use of the rate of naturally occurring behaviors as applied in some *G* studies (e.g., **Huber**; **Shuker et al.** See also Reader et al. [2011] for a discussion of the advantages and disadvantages of this procedure). However, if this method would only produce noise, one would not find strong correlations with brain size or aspects of executive functions, or that *G* is the principal locus of selection in the macroevolution of primate intelligence (Fernandes et al. 2014). Moreover, the anomaly of the lack of success of the mixed intraspecific and interspecific studies to generate a common main factor has been solved by **Woodley of Menie, Fernandes, te Nijenhuis, Aguirre, & Figueredo**. They suggested that variables with floor or ceiling effects may obscure differences in general intelligence across species because they cannot load on *g*. Their analysis supports this idea because species differences are especially striking for tests that load highly on *g*. Overall, then, the increasing plausibility of the idea that *g* and *G* can be equated automatically supports the argument that animals have something that closely resembles human *g*, and may even be homologous to it. This position would also be consistent with the increasing knowledge of the nature of

cortical development and especially plastic responses to external inputs during brain development (Anderson & Finlay 2014). It also should serve to refute the alternative interpretations of *G* as totally unrelated to *g*, offered by **Arden & Zietsch** and **Lewis et al.**

In order to resolve current debates, we believe empirical work is needed that confirms the presence of domain-general processes. First, if *g* were all about health, myelinization, or developmental stability – that is, non-cognitive features – we would expect a correlation between all cognitive abilities, regardless of whether they derive from primary or secondary modules. But if we were dealing with general intelligence, the pairwise correlations between abilities deriving from primary modules should be weaker (largely because they should all be at ceiling or floor values).

Second, more specific tests can be done to provide external validation for the interpretation of domain-general intelligence, along the lines outlined in the target article (sect. 2.5, Table 7) and as also stressed by **Locurto**. We can derive additional ones from the commentators' comments. **Arslan et al.** propose to use genetically uniform strains and mutation accumulation lines to help clarify the role of genetic load, and **Lee & Chabris** propose to use genome-wide association studies to test if distinct cognitive abilities are genetically correlated and thus potentially the target of natural selection. More specifically, **Pennycook & Rand-** suggest to focus on meta-cognitive abilities; **van Horik & Lea** stress the usefulness of reversal learning paradigms; and **Jacobs & Gärdenfors**, causal reasoning. **Pepperberg** summarizes additional fruitful paradigms in which several bird species excel. **Bar-Hen-Schweiger et al.** propose to also include object manipulation. This is intriguing because recent results show that object manipulation complexity is indeed correlated with brain size (Heldstab et al. 2016), but to what extent mental manipulation can be seen as a direct extension of object manipulation requires additional research, in particular in nonhuman animals. For interspecific studies, species differences in basal ganglia–based circuitry may provide further insight into the evolution of *G*, as highlighted by **Nordli & Todd**.

In sum, we acknowledge and agree that the mere presence of a first PCA factor (*g*) in intraspecific psychometric studies is not sufficient evidence for general intelligence, but that such an interpretation critically requires evidence for predictive or criterion-related validity as stressed by **Locurto** and outlined in section 2.5 in the target article. However, we also argue that if a *g* factor becomes manifest in psychometric studies, and if this *g* is not an artifact and shows predictive and criterion-related validity, how exactly the positive intercorrelations arise no longer matters for the claim that *g* was the target of selection. In other words, it may well be that *g* is an emergent property of the central neural system, and we in fact think it is unlikely that it can be reduced to any specific psychological or biological trait or construct, and thus agree on this with **Ramus** or **Arden & Zietsch**. But this does not imply that *g* as potentially emergent property cannot be the target of selection. This view is supported by the majority of results from interspecific *G* studies.

R6. *g*/*G* and brain size

Brain size is reliably associated with *G* in interspecific primate studies, and less strongly with *g* in intraspecific studies. **Matzel & Sauce** argue that brain size cannot be important in intelligence. First, they argue that Neanderthals had larger brains than humans, but were "cognitively inferior" to modern humans. However, cultural intelligence can explain this difference because modern humans had much larger groups (Mellars & French 2011), and social inputs play a crucial role in affecting the size of a population's skill repertoires (see also Henrich 2016). Thus, differences in brain size (provided they also extended to differences in relative brain size) can be compensated for by differences in the quality and quantity of inputs during development. This observation also deals with their second objection, namely that brain size predicts only a modest proportion of variation in human *g*.

The second argument by **Matzel & Sauce** as to why brain size should not matter for intelligence was that intelligence can look quite similar in lineages with very different absolute or relative brain sizes, particularly in birds versus mammals. This can partly be explained by major between-lineage differences in neuron densities, between cetaceans and other mammals, and between mammals and birds (Olkowicz et al. 2016). Accordingly, comparative studies on relative brain size within birds are consistent with the presence of general intelligence in this lineage (Lefebvre et al. 2004). Nonetheless, it would be useful to examine the degree to which more extensive tests of general intelligence in birds (e.g., Isden et al. 2013; Shaw et al. 2015) show the same structure of cognition as in humans and presumably other mammals. If multiple independent origins can indeed be demonstrated, as suggested by **Matzel & Sauce**, this would support the contention that beyond a certain level of complexity, domain-general processes evolve to take over the control of procedures that serve a useful function in many different domains (**Wang**).

R7. *g* and biological fitness

If general intelligence is an adaptation, then we expect some link to fitness. Thus, one intuitive way of assessing predictive validity of *g* is to ask if individuals scoring higher on *g* have better survival and higher reproductive success. This basic idea is compelling, but both empirical and conceptual caveats need to be carefully considered.

In humans, reproductive success is often no longer maximized in modern societies due to the demographic transition (Coale 1989). Survival, however, continues to be a high individual priority and, therefore, is a better measure for this link in humans. There is in fact extensive evidence that *g* scores predict survival or longevity (Deary 2008).

If general intelligence is an adaptation in nonhuman species, too, then the estimates of this ability should also correlate with fitness measures. This correlation has in fact been proposed as a test of the predictive validity of the concept (**Locurto**). Tests in the wild presuppose that we can actually estimate general intelligence in wild animals. This may be feasible in birds (Isden et al. 2013; Shaw et al. 2015), but will be extremely difficult in many other species that show strong neophobia in the wild (Forss et al. 2015). Moreover, as stressed by **van Horik & Lea**, it is more difficult to control for confounds or selective participation, or to retest the same individuals (see also Rowe & Healy 2014).

Yet, we must also be careful for conceptual reasons. General intelligence is developmentally constructed, and investment in increased intelligence may trade off against other vital activities. It is, therefore, quite conceivable that estimates of general intelligence need not show a positive relationship with fitness but, rather, an inverted U-shaped relationship. Thus, if investment in general intelligence and the brain requires resources that also positively affect other vital processes such as growth and reproduction, this can have a negative impact on an individual's fitness. The positive relationship in modern humans may thus be something of an anomaly, perhaps linked to the absence of such tradeoffs (although it could well be that it existed for reproduction). Indeed, the emerging work examining the link between cognitive abilities and fitness (Morand-Ferron & Quinn 2015) produces mixed results. One reason for this may be that a direct positive relationship with fitness is expected for primary modules, but not for secondary modules. This topic is worth exploring in greater depth if practical obstacles can be overcome (see also Morand-Ferron et al. 2015; Rowe & Healy 2014).

R8. Cultural intelligence

The presence of high general intelligence in animals poses something of a conundrum. Individual problem solving based purely on general cognitive ability will only rarely produce useful behavioral innovations, but will often also carry risks of injury or poisoning. Moreover, rarely invented innovations are not transmitted to the next generation; only the ability to make such inventions is. The cultural intelligence hypothesis, therefore, suggests that general intelligence is most likely to evolve where the process of problem solving is socially canalized – that is, in species that strongly rely on social learning, because social canalization increases the odds of actually finding a solution. It stresses that social information can efficiently guide exploration during development. It is, therefore, as **Whiten** correctly points out, an enabling hypothesis, which concerns the acquisition process of cognitive skills (and the conditions selecting for them), but is silent on the actual nature of the cognitive benefits. The latter could often be cognitive flexibility serving as a cognitive buffer, allowing animals to survive in seasonal habitats, as suggested by **Holekamp & Miikkulainen** (cf. van Woerden et al. 2012). However, this functional benefit should have been favored in many species, and the cultural intelligence hypothesis argues that selection is more likely to produce this adaptive benefit for behavioral flexibility where costs remain low due to many opportunities to acquire social information.

Cognition requires brains, and brains are energetically expensive. Thus, the significant increase in energy need accompanying increased brain size must be met, directly or indirectly. As shown in Figure R2, natural selection would, therefore, favor those mechanisms that either provide a direct energy boost (for instance, ecological effects of cognition leading to reduced fluctuations in food intake: arrow 1) or allow for reduced allocation to growth and reproduction due to improved survival (arrow 2). The far greater efficiency provided by skill acquisition through social learning rather than inefficient and risky innovation (arrow 3) is another selective benefit highlighted

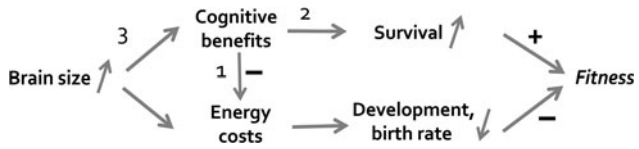


Figure R2. Processes that could be affected by selection for increased brain size: (1) an immediate effect of improved cognition on reduction in energy costs; (2) an unusually large fitness benefit from possessing cognitive skills; or (3) an unusually efficient translation of brain tissue into cognitive skills, as argued by the cultural intelligence hypothesis. The argument may hold especially for domain-general processes, because the latter appear to require exceptional amounts of brain tissue.

by the cultural intelligence hypothesis. This argument clearly implies, although we did not point this out explicitly, that the cultural intelligence hypothesis is, like almost all evolutionary hypotheses, a *ceteris paribus* hypothesis: Holding all other conditions constant, the availability of social information should make it easier to evolve the supporting infrastructure for increased behavioral flexibility in a particular lineage.

There may also be other conditions favoring the evolution of intelligence. Cephalopods, especially octopus, have no contact between generations, and mustelids disperse at a very young age and then remain solitary, yet show signs of strong cognitive performance (**Holekamp & Miikkulainen**; cf. Holekamp et al. 2015). In the expensive brain framework, this would suggest unusual payoffs from exploration and innovation (arrow 2 in Figure R2), or unusually low risks of exploration – which are testable predictions. These seemingly intelligent solitary species, therefore, would appear to be exceptions to the cultural intelligence hypothesis that should be acknowledged, although the impressive imitation ability of solitary octopus (Fiorito & Scotto 1992) is puzzling. The possibility of alternative evolutionary pathways to intelligence makes it even more important to explicitly deploy extensive test batteries to examine if these species truly show evidence of general intelligence, or whether we find specialized (albeit quite impressive) domain-specific cognitive adaptations.

R9. Conclusions

Our goal is to understand the evolutionary origin of general intelligence. The case is still open, even though the body of evidence suggests we also find it in other species and that the cultural intelligence approach can help us understand why it arose in some species but not in others. But identifying the evolutionary origin of *g* or *G* in nonhumans is not the full answer to how the human mind evolved. Humans are more than *g* (**Amici et al.**; **Gardner**; **Pennycook & Rand**) and other components are important too, first and foremost language, but also components summarized as multiple intelligences by **Gardner**. In principle, the same approach as taken here can be applied to each of these components, by asking what the interspecific distribution of this trait is, and what factors may best explain this distribution. For general intelligence, the most likely factor is brain size, whereas for prosocial attitude, allomaternal care seems to explain interspecific distribution (Burkart et al. 2014).

Cultural intelligence effects are particularly pronounced in humans. Among others, this is evident in the Flynn effect, and the strong influence of environment on general intelligence, particularly early in ontogeny when individuals still cannot seek out an environment that matches their abilities (Flynn 2016). Human cultural intelligence is supported by additional adaptations such as language and pedagogy, which involves strong theory-of-mind abilities and intentional teaching as pointed out by **Estes & Bartsch** (see also Burkart et al. 2009). For humans in particular, one would expect, therefore, that socio-cognitive abilities predict general intelligence, and **Fernández-Berrocal, Cabello, & Gutiérrez-Cobo** present evidence for humans supporting this prediction (for interspecific tests of this link, **Arslan et al.** suggest joint phylogenetic tree analyses of sociality and variance explained by *g*, but this will not work because the amount of variance explained is not a good measure for how much "*g*" a given species "has," as detailed in the target article (sect. 2.4.1)).

References

[The letters "a" and "r" before author's initials stand for target article and response references, respectively]

Abutalebi, J. & Clahsen, H. (2015) Bilingualism, cognition, and aging. *Bilingualism: Language and Cognition* 18(01):1–2. [aJMB]

Aiello, L. C. & Dean, M. C. (1990) *An introduction to human evolutionary anatomy*. Academic Press. [LDM]

Aldridge, J. W., Berridge, K. C. & Rosen, A. R. (2004) Basal ganglia neural mechanisms of natural movement sequences. *Canadian Journal of Physiology and Pharmacology* 82:732–39. doi: 10.1139/Y04-061. [SAN]

Allman, J., McLaughlin, T. & Hakeem, A. (1993) Brain weight and life-span in primate species. *Proceedings of the National Academy of Sciences USA* 90:118–22. [KEH]

Amici, F., Aureli, F. & Call, J. (2008) Fission-fusion dynamics, behavioral flexibility, and inhibitory control in primates. *Current Biology* 18(18):1415–19. [aJMB]

Amici, F., Aureli, F. & Call, J. (2010) Monkeys and apes: Are their cognitive skills really so different? *American Journal of Physical Anthropology* 143(2):188–97. [aJMB]

Amici, F., Barney, B., Johnson, V. E., Call, J. & Aureli, F. (2012) A modular mind? A test using individual data from seven primate species. *PLoS One* 7(12):e51918. [aJMB, FA]

Amiel, J. J., Tingley, R. & Shine, R. (2011) Smart moves: Effects of relative brain size on establishment success of invasive amphibians and reptiles. *PLoS One* 6(4): e18277. [aJMB]

Aminoff, E., Gronau, N. & Bar, M. (2007) The parahippocampal cortex mediates spatial and nonspatial associations. *Cerebral Cortex* 17(7):1493–503. Available at: http://doi.org/10.1093/cercor/bhl078. [MdJDM]

Anastasi, A. (1961) *Psychological testing* (3rd edition). Macmillan. [CL]

Anderson, B. (1993) Evidence from the rat for a general factor that underlies cognitive performance and that relates to brain size: Intelligence? *Neuroscience Letters* 153(1):98–102. [aJMB, LH]

Anderson, M. L. (2010) Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences* 33:245–313. [MB-H-S]

Anderson, M. L. (2016) Précis of *After Phrenology: Neural Reuse and the Interactive Brain*. *Behavioral and Brain Sciences* 39:e120. doi: 10.1017/S0140525X15000631. [MdJDM]

Anderson, M. L. & Finlay, B. L. (2014) Allocating structure to function: The strong links between neuroplasticity and natural selection. *Frontiers in Human Neuroscience* 7:918, 1–16. [arJMB]

Anderson, R. C., Mather, J. A., Monette, M. Q. & Zimsen, S. R. M. (2010) Octopuses (Enteroctopus dofleini) recognize individual humans. *Journal of Applied Animal Welfare Science* 13:261–72. [KEH]

Aplin, L. M., Farine, D. R., Morand-Ferron, J., Cockburn, A., Thornton, A. & Sheldon, B. C. (2015) Experimentally induced innovations lead to persistent culture via conformity in wild birds. *Nature* 518(7540):538–41. [aJMB]

Arden, R. & Adams, M. J. (2016) A general intelligence factor in dogs. *Intelligence* 55:79–85. [aJMB]

Arden, R., Bensky, M. K. & Adams, M. J. (2016) A review of cognitive abilities in dogs, 1911 through 2016: More individual differences, please! *Current Directions in Psychological Science* 25:307–12. doi: 10.1177/0963721416667718. [RCA]

Arslan, R. C. & Penke, L. (2015) Evolutionary genetics. In: *Handbook of evolutionary psychology*, vol. 2, ed. D. M. Buss, pp. 1047–66. Wiley. [RCA]

Auersperg, A. M. I., Laumer, I. B. & Bugnyar, T. (2013) Goffin cockatoos wait for qualitative and quantitative gains but prefer "better" to "more". *The Royal Society: Biology Letters* 9 (article 20121092). [IMP]

Aust, U., Range, F., Steurer, M. & Huber, L. (2008) Inferential reasoning by exclusion in pigeons, dogs, and humans. *Animal Cognition* 11:587–97. [LH]

Baddeley, A. (2010) Working memory. *Current Biology* 20:R136–40. [aJMB]

Bagozzi, R. P. (2007) On the meaning of formative measurement and how it differs from reflective measurement: Comment on Howell, Breivik, and Wilcox (2007). *Psychological Methods* 12(2):229–37; discussion 238–45. doi: 10.1037/1082-989X.12.2.229. [KK]

Bailey, A. M., McDaniel, W. F. & Thomas, R. K. (2007) Approaches to the study of higher cognitive functions related to creativity in nonhuman animals. *Methods* 42(1):3–11. [aJMB]

Baker, C. A., Peterson, E., Pulos, S. & Kirkland, R. A. (2014) Eyes and IQ: A meta-analysis of the relationship between intelligence and "Reading the Mind in the Eyes". *Intelligence* 44:78–92. [PF-B]

Banerjee, K., Chabris, C. F., Johnson, V. E., Lee, J. J., Tsao, F. & Hauser, M. D. (2009) General intelligence in another primate: Individual differences across cognitive task performance in a New World monkey (Saguinus oedipus). *PLoS ONE* 4(6):e5883. [aJMB]

Banks, W. E., d'Errico, F., Peterson, A. T., Kageyama, M., Sima, A. & Sanchez-Goni, M. F. (2008) Neanderthal extinction by competitive exclusion. *PLoS One* 3(12): e3972. doi: 10.1371/journal.pone.0003972. [LDM]

Baptista, L. F. (1981) Interspecific song mimesis by a Lincoln sparrow. *Wilson Bulletin* 93:265–67. [IMP]

Barbey, A. K., Colom, R., Solomon, J., Krueger, F., Forbes, C. & Grafman, J. (2012) An integrative architecture for general intelligence and executive function revealed by lesion mapping. *Brain* 135(4):1154–64. [aJMB]

Barnes, T. D., Kubota, Y., Hu, D., Jin, D. Z. & Graybiel, A. M. (2005) Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature* 437:1158–61. doi: 10.1038/nature04053. [SAN]

Barney, B. J., Amici, F., Aureli, F., Call, J. & Johnson, V. E. (2015) Joint Bayesian modeling of binomial and rank data for primate cognition. *Journal of the American Statistical Association* 110(510):573–82. [aJMB]

Baron-Cohen, S., Jolliffe, T., Mortimore, C. & Robertson, M. (1997) Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger's syndrome. *Journal of Child Psychology and Psychiatry* 38:813–22. [PF-B]

Baron-Cohen, S., Leslie, A. M. & Frith, U. (1985) Does the autistic child have a "theory of mind"? *Cognition* 21:37–46. [PF-B]

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y. & Plumb, I. (2001) The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry* 42:241–51. [PF-B]

Barrett, H. C. (2015) Modularity. In: *Evolutionary perspectives on social psychology*, ed. V. Zeigler-Hill, L. L. M. Welling & T. K. Schackelford, pp. 39–51. Springer International. [arJMB]

Barrett, H. C. & Kurzban, R. (2006) Modularity in cognition: Framing the debate. *Psychological Review* 113(3):628–47. [aJMB]

Barrett, H. C. & Kurzban, R. (2012) What are the functions of System 2 modules? A reply to Chiappe and Gardner. *Theory & Psychology* 22(5):683–88. [aJMB]

Barrett, L. (2011) *Beyond the brain: How body and environment shape animal and human minds*. Princeton University Press. [RA]

Bartholomew, D. J., Deary, I. J. & Lawn, M. (2009) A new lease of life for Thomson's bonds model of intelligence. *Psychological Review* 116:567–79. Available at: https://doi.org/10.1037/a0016262. [aJMB, FR]

Barton, R. A. (2012) Embodied cognitive evolution and the cerebellum. *Philosophical Transactions of the Royal Society* 367:2097–107. [DS]

Barton, R. A. & Harvey, P. H. (2000) Mosaic evolution of brain structure in mammals. *Nature* 405:1055–58. [FA]

Bartsch, K., Horvath, K. & Estes, D. (2003) Young children's talk about learning events. *Cognitive Development* 18:177–93. [DE]

Bateson, P. & Curley, J. (2013) Developmental approaches to behavioural biology. *Nova Acta Leopoldina* 111:89–110. [IJ]

Bear, A. & Rand, D. G. (2016a) Intuition, deliberation, and the evolution of cooperation. *Proceedings of the National Academy of Sciences USA* 113:1–6. doi: 10.1073/pnas.1517780113. [GP]

Bear, A. & Rand, D. G. (2016b) Modeling intuition's origins. *Journal of Applied Memory & Cognition* 5:341–44. [GP]

Bear, A., Kagan, A. & Rand, D. G. (2016) Co-evolution of cooperation and cognition: The impact of imperfect deliberation and context-sensitive intuition. (November 1, 2016). Available at SSRN: https://ssrn.com/abstract=2862717. [GP]

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W. & Rushworth, M. F. S. (2008) Associative learning of social value. *Nature* 456(7219):245–49. [aJMB]

Bialystok, E., Craik, F. I. & Luk, G. (2012) Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences* 16(4):240–50. [aJMB]

Bilalić, M., Langner, R., Ulrich, R. & Grodd, W. (2011) Many faces of expertise: Fusiform face area in chess experts and novices. *The Journal of Neuroscience* 31 (28):10206–14. [aJMB]

Bitterman, M. E. (1960) Toward a comparative psychology of learning. *American Psychologist* 15:704–12. [RJS]

Blair, C. (2006) How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences* 29(2):109–25; discussion 125–60. doi: 10.1017/S0140525X06009034. [aJMB, KK]

Blaisdell, A. P., Sawa, K., Leising, K. J. & Waldmann, M. R. (2006) Causal reasoning in rats. *Science* 311:1020–22. [IJ]

Bolhuis, J. J. (1991) Mechanisms of avian imprinting: A review. *Biological Reviews* 66:303–45. doi: 10.1111/j.1469-185X.1991.tb01145.x. [IJ]

Bolhuis, J. J., Brown, G. R., Richardson, R. C. & Laland, K. N. (2011) Darwin in mind: New opportunities for evolutionary psychology. *PLoS Biology* 9(7): e1001109. [aJMB]

Borsboom, D. & Dolan, C. V. (2006) Why *g* is not an adaptation: A comment on Kanazawa (2004) *Psychological Review* 113(2):433–37. Available at: http://doi.org/10.1037/0033-295X.113.2.433. [RA]

Bouchard, T. J. (2014) Genes, evolution and intelligence. *Behavioural Genetics* 44 (6):549–77. [aJMB]

Box, H. O. (1984) *Primate behavior and social ecology*. Chapman and Hall. [aJMB]

Boyd, R., Richerson, P. J. & Henrich, J. (2011) The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences USA* 108(Supplement 2):10918–25. [aJMB]

Bradford, E. E. F., Jentzsch, I. & Gomez, J.-C. (2015) From self to social cognition: Theory of mind mechanisms and their relation to executive functioning. *Cognition* 138:21–34. [MDH]

Brodin, A. (2010) The history of scatter hoarding studies. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365(1542):869–81. [aJMB]

Brooks, R. A. (1990) Elephants don't play chess. *Robotics and Autonomous Systems* 6:1–16. doi: 10.1016/S0921-8890(05)80024-7. [KEH]

Brooks, R. A. (1991) Intelligence without reason. In: *Proceedings of the 12th International Joint Conference on Artificial Intelligence* (IJCAI-91), pp. 569–95. Morgan Kaufmann. [KEH]

Brown, J., Kaplan, G., Rogers, L. J. & Vallortigara, G. (2010) Perception of biological motion in common marmosets (*Callithrix jacchus*): By females only. *Animal Cognition* 13(3):555–64. [aJMB]

Browne, M., Thomson, P., Rockloff, M. & Pennycook, G. (2015) Going against the herd: Understanding the psychosocial factors underlying the "vaccination confidence gap". *PLoS One* 10:e1032562. [GP]

Bugnyar, T. & Heinrich, B. (2006) Pilfering ravens, *Corvus corax*, adjust their behaviour to social context and identity of competitors. *Animal Cognition* 9:369–76. [IMP]

Burgess, G. C., Gray, J. R., Conway, A. R. A. & Braver, T. S. (2011) Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology: General* 140(4):674. [aJMB]

Burkart, J. M. (2017) The evolution and consequences of sociality. In: *Oxford handbook of comparative psychology*, ed. J. Call, G. M. Burghardt, I. M. Pepperberg, C. T. Snowdon & T. Zentall. Oxford University Press. [aJMB]

Burkart, J. M., Allon, O., Amici, F., Fichtel, C., Finkenwirth, C., Heschl, A., Huber, J., Isler, K., Kosonen, Z., Martins, E., Meulman, E., Richiger, R., Rueth, K., Spillmann, B., Wiesendanger, S. & van Schaik, C. P. (2014) The evolutionary origin of human hyper-cooperation. *Nature Communications* 5:4747. [rJMB]

Burkart, J. M., Hrdy, S. B. & van Schaik, C. P. (2009) Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology* 18:175–86. [arJMB]

Burkart, J. M. & van Schaik, C. P. (2010) Cognitive consequences of cooperative breeding. *Animal Cognition* 31:1–19. [aJMB]

Burkart, J. M. & van Schaik, C. P. (2016a) Revisiting the consequences of cooperative breeding. A response to Thornton & McAuliffe (2015) *Journal of Zoology* 299:77–83. doi:10.1111/jzo.12322. [aJMB]

Burkart, J. M. & van Schaik, C. P. (2016b) The cooperative breeding perspective helps pinning down when uniquely human evolutionary processes are necessary. Commentary to Richerson et al. (2016) *Behavioral and Brain Sciences* 39: e34. [aJMB]

Byrne, R. W. (1994) The evolution of intelligence. In: *Behaviour and evolution*, ed. P. J. B. Slater & T. R. Halliday, Cambridge University Press. [aJMB]

Byrne, R. W. & Corp, N. (2004) Neocortex size predicts deception rate in primates. *Proceedings of the Royal Society of London B: Biological Sciences* 271:1693–99. [aJMB, MB-H-S]

Byrne, R. W. & Whiten, A. (1990) Tactical deception in primates: The 1990 database. *Primate Report* 27:1–101. [aJMB]

Cabello, R., Sorrel, M. A., Fernández-Pinto, I., Extremera, N. & Fernández-Berrocal, P. (2016) Age and gender differences in ability emotional intelligence in adults: A cross-sectional study. *Developmental Psychology* 52:1486–92. [PF-B]

Cardoso, R. M. (2013) Resolução de problema por macacos-prego selvagens (Sapajus libidinosus) de duas populações com diferentes repertórios de uso de ferramentas. Ph.D. dissertation, Departamento de Psicologia Experimental, Universidade de São Paulo. [aJMB]

Carey, S. (2009) *The origin of concepts*. Oxford University Press. [aJMB]

Carpenter, M., Nagell, K. & Tomasello, M. (1998) Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development* 63(4, Serial No. 255). [DE]

Carroll, J. B. (1993) *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. [aJMB]

Carruthers, P. (2005) The case for massively modular models of mind. In: *Contemporary debates in cognitive science*, ed. R. Stainton, pp. 205–25. Blackwell. [aJMB]

Carruthers, P. (2011) I do not exist. *Trends in Cognitive Sciences* 15(5):189–90. [aJMB]

Catani, M., Jones, D. K. & Ffytche, D. H. (2005) Perisylvian language networks of the human brain. *Annals of Neurology* 57(1):8–16. Available at: http://doi.org/10.1002/ana.20319. [MdJDM]

Cattell, R. B. (1963) Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology* 54(1):1–22. [aJMB]

Cattell, R. B. (1971) *Abilities: Their structure, growth, and action*. Houghton Mifflin. Available at: http://books.google.com/books?id=10EgAQAAIAAJ&pgis=1. [KK]

Cattell, R. B. (1987) *Intelligence: Its structure, growth and action*. Elsevier. [aJMB]

Cauchoix, M. & Chaine, A. S. (2016) How can we study the evolution of animal minds? *Frontiers in Psychology* 7:358. [rJMB]

Chabris, C. F. (2007) Cognitive and neurobiological mechanisms of the law of general intelligence. In: *Integrating the mind: Domain general versus domain specific processes in higher cognition*, ed. M. J. Roberts, pp. 449–91. Psychology Press. [aJMB]

Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. B. & Laibson, D. I. (2015) The fourth law of behavior genetics. *Current Directions in Psychological Science* 24:304–12. [JJL]

Chang, Y. (2014) Reorganization and plastic changes of the human brain associated with skill learning and expertise. *Frontiers in Human Neuroscience* 8:35. [aJMB]

Chiappe, D. & Gardner, R. A. (2012) The modularity debate in evolutionary psychology. *Theory & Psychology* 22(5):669–82. [aJMB]

Chiarati, E., Canestrari, D., Vera, R., Marcos, J. M. & Baglione, V. (2010) Linear and stable dominance hierarchies in cooperative carrion crows. *Ethology* 116:346–56. [IMP]

Chittka, L. & Niven, J. (2009) Are bigger brains better? *Current Biology* 19(21): R995–1008. [aJMB]

Chudasama, Y. (2011) Animal models of prefrontal-executive function. *Behavioral Neuroscience* 125(3):327–43. [aJMB]

Clayton, N. S. (2001) Hippocampal growth and maintenance depend on food-caching experience in juvenile mountain chickadees (Poecile gambeli). *Behavioral Neuroscience.* 115(3):614–25. [LDM]

Clune, J., Mouret, J. B. & Lipson, H. (2013) The evolutionary origins of modularity. *Proceedings of the Royal Society of London B: Biological Sciences* 280 (1755):20122863. [aJMB]

Clutton-Brock, T. H. & Harvey, P. H. (1977) Primate ecology and social organization. *Journal of Zoology* 183:1–39. [DS]

Coale, A. J. (1989) Demographic transition. In: *Social economics*, ed. Eatwell, J., Milgate, M. & Newman, P. pp. 16–23. Palgrave Macmillan. [rJMB]

Coehlo, C. G., Falotico, T., Ozar, P., Mannu, M., Resende, B. D., Siqueira, J. O. & Otonni, E. (2015) Social learning strategies for nut-cracking by tufted capuchin monkeys (*Sapajus* spp.). *Animal Cognition* 18:911–19. [AW]

Cole, M., Gay, J., Glick, J. & Sharp, D. W. (1971) *The cultural context of learning and thinking*. Basic Books. [RJS]

Colom, R., Abad, F. J., Rebollo, I. & Shih, P. C. (2005) Memory span and general intelligence: A latent-variable approach. *Intelligence* 33:623–42. [aJMB]

Colom, R., Jung, R. E. & Haier, R. J. (2006) Distributed brain sites for the g-factor of intelligence. *NeuroImage* 31(3):1359–65. [aJMB]

Colom, R., Román, F. J., Abad, F. J., Shih, P. C., Privado, J., Froufe, M., Escorial, S., Martínez, K., Burgaleta, M., Quiroga, M. A., Karama, S., Haier, R. J., Thompson, P. M. & Jaeggi, S. M. (2013) Adaptive n-back training does not improve fluid intelligence at the construct level: Gains on individual tests suggest that training may enhance visuospatial processing. *Intelligence* 41:712–27. doi: 10.1016/j.intell.2013.09.002. [RCA]

Coltheart, M. (2011) Methods for modular modelling: Additive factors and cognitive neuropsychology. *Cognitive Neuropsychology* 28(3–4):224–40. [aJMB]

Confer, J. C., Easton, J. A., Fleischman, D. S., Goetz, C. D., Lewis, D. M. G., Perilloux, C. & Buss, D. M. (2010) Evolutionary psychology: Controversies, questions, prospects, and limitations. *American Psychologist* 65:110–26. [DMGL]

Conway, A. R. A. & Kovacs, K. (2013) Individual differences in intelligence and working memory: A review of latent variable models. In: *Psychology of Learning and Motivation, vol. 58*, ed. B. H. Ross, pp. 233–70. Academic Press. doi: 10.1016/B978-0-12-407237-4.00007-4. [KK]

Cook, M. & Mineka, S. (1989) Observational conditioning of fear to fear-relevant versus fear-irrelevant stimuli in rhesus monkeys. *Journal of Abnormal Psychology* 98:448–59. [aJMB]

Cosmides, L. & Tooby, J. (1994) Origins of domain specificity: The evolution of functional organization. In: *Mapping the mind: Domain specificity in cognition and culture*, ed. L. Hirschfeld & S. Gelman, pp. 85–116. Cambridge University Press. [aJMB]

Cosmides, L. & Tooby, J. (2002) Unraveling the enigma of human intelligence: Evolutionary psychology and the multimodular mind. In: *The evolution of intelligence*, ed. R. J. Sternberg & J. C. Kaufman, pp. 145–98. Erlbaum. [aJMB, FA]

Cosmides, L. & Tooby, J. (2013) Evolutionary psychology: New perspectives on cognition and motivation. *Annual Review of Psychology* 64:201–29. [aJMB, X-TW]

Cosmides, L., Barrett, H. C. & Tooby, J. (2010) Adaptive specializations, social exchange, and the evolution of human intelligence. *Proceedings of the National Academy of Sciences USA* 107(Supplement 2):9007–14. [aJMB, X-TW]

Csibra, G. & Gergely, G. (2009) Natural pedagogy. *Trends in Cognitive Sciences* 13:148–53. doi: 10.1016/j.tics.2009.01.005. [DE]

d'Souza, D. & Karmiloff-Smith, A. (2011) When modularization fails to occur: A developmental perspective. *Cognitive Neuropsychology* 28(3–4):276–87. [aJMB]

Dahl, C. D., Chen, C. C. & Rasch, M. J. (2014) Own-race and own-species advantages in face perception: A computational view. *Scientific Reports* 4:6654. [aJMB]

Danforth, J. S., Chase, P. N., Dolan, M. & Joyce, J. H. (1990) The establishment of stimulus-control by instructions and by differential reinforcement. *Journal of the Experimental Analysis of Behavior* 54:97–112. doi: 10.1901/jeab.1990.54-97. [JOvH]

Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D., Liewald D, Ke X, Le Hellard, S., Christoforou, A., Luciano, M., McGhee, K., Lopez, L., Gow, A. J., Corley, J., Redmond, P., Fox, H. C., Haggarty, P., Whalley, L. J., McNeill, G., Goddard, M. E., Espeseth, T., Lundervold, A. J., Reinvang, I., Pickles, A., Steen, V. M., Ollier, W., Porteous, D. J., Horan, M., Starr, J. M., Pendleton, N., Visscher, P. M. & Deary, I. J. (2011) Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry* 16:996–1005. [aJMB]

de Waal, F. B. M. & Johanowicz, D. L. (1993) Modification of reconciliation behavior through social experience: An experiment with two macaque species. *Child Development* 64:897–908. [FA]

de Winter, W. & Oxnard, C. E. (2001) Evolutionary radiations and convergences in the structural organization of mammalian brains. *Nature* 409:710–14. [FA]

Dean, L. G., Vale, G. L., Laland, K. N., Flynn, E. & Kendal, R. L. (2014) Human cumulative culture: A comparative perspective. *Biological Reviews* 89(2):284–301. [aJMB]

Deaner, R. O., Barton, R. A. & van Schaik, C. P. (2003) Primate brains and life history: Renewing the connection. In: *Primate Life Histories and Socioecology*, eds. P. M. Kappeler & M. E. Pereira, pp. 233–65. University of Chicago Press. [KEH]

Deaner, R. O., Isler, K., Burkart, J. M. & van Schaik, C. P. (2007) Overall brain size, and not encephalization quotient, best predicts cognitive ability across non-human primates. *Brain, Behaviour and Evolution* 70(2):115–24. [aJMB]

Deaner, R. O., van Schaik, C. P. & Johnson, V. E. (2006) Do some taxa have better domain-general cognition than others? A meta-analysis of nonhuman primate studies. *Evolutionary Psychology* 4:149–96. [aJMB]

Deary, I. J. (2008) Why do intelligent people live longer? *Nature* 456(7219):175–76. [rJMB]

Deary, I. J. (2012) Looking for "system integrity" in cognitive epidemiology. *Gerontology* 58:545–53. doi: 10.1159/000341157. [RCA]

Deary, I. J., Cox, S. R. & Ritchie, S. J. (2016) Getting Spearman off the skyhook: One more in a century (since Thomson, 1916) of attempts to vanquish *g*. *Psychological Inquiry* 27(3):192–99. [rJMB]

Deary, I. J., Penke, L. & Johnson, W. (2010) The neuroscience of human intelligence differences. *Nature Reviews Neuroscience* 11(3):201–11. [arJMB, MB-H-S]

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L. & Pallier, C. (2015) The Neural representation of sequences: From transition probabilities to algebraic patterns and linguistic trees. *Neuron* 88(1):2–19. Available at: http://doi.org/10.1016/j.neuron.2015.09.019. [MdJDM]

Detterman, D. (1993) *Individual differences and cognition*. Praeger. [HG]

Diamond, A. (2013) Executive functions. *Annual Review of Psychology* 64:135. [aJMB]

Diekamp, B., Kalt, T. & Gunturkun, O. (2002) Working memory neurons in pigeons. *Journal of Neuroscience* 22(4):RC210. [LDM]

Dong, X., Wang, X., Zhang, F. & Tian, W. (2016) Genome-wide identification of regulatory sequences undergoing accelerated evolution in the human genome. *Molecular Biology and Evolution* 33:2565–75. [JJL]

Drake, J. M. (2007) Parental investment and fecundity, but not brain size, are associated with establishment success in introduced fishes. *Functional Ecology* 21(5):963–68. [aJMB]

Ducatez, S., Clavel, J. & Lefebvre, L. (2015) Ecological generalism and behavioural innovation in birds: Technical intelligence or the simple incorporation of new foods? *Journal of Animal Ecology* 84(1):79–89. [aJMB]

Duchaine, B., Cosmides, L. & Tooby, J. (2001) Evolutionary psychology and the brain. *Current Opinion in Neurobiology* 11(2):225–30. [aJMB]

Dunbar, R. (1998) Theory of mind and the evolution of language. In: *Approaches to the evolution of language*, ed. J. Hurford, M. Studdert-Kennedy & C. Knight, pp. 92–110. Cambridge University Press. [DE]

Dunbar, R. I. M. (1992) Neocortex size as a constraint on group size in primates. *Journal of Human Evolution* 20:469–93. [aJMB]

Dunbar, R. I. M. & Shultz, S. (2007a) Evolution in the social brain. *Science* 317:1344–47. [aJMB, AW]

Dunbar, R. I. M. & Shultz, S. (2007b) Understanding primate brain evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* 362 (1480):649–58. [aJMB]

Duncan, J., Burgess, P. & Emslie, H. (1995) Fluid intelligence after frontal lobe lesions. *Neuropsychologia* 33(3):261–68. doi: 10.1016/0028-3932(94)00124-8. [KK]

Embretson, S. E. (1995) The role of working memory capacity and general control processes in intelligence. *Intelligence* 20(2):169–89. [aJMB]

Emery, N. J. & Clayton, N. S. (2004) The mentality of crows: Convergent evolution of intelligence in corvids and apes. *Science* 306(5703):1903–907. [LH]

Emery, N. J. & Clayton, N. S. (2009) Tool use and physical cognition in birds and mammals. *Current Opinion in Neurobiology* 19(1):27–33. [LH]

Enard, W., Gehre, S., Hammerschmidt, K., Hölter, S. M., Blass, T., Somel, M., Brückner, M. K., Schreiweis, C., Winter, C., Sohr, R., Becker, L., Wiebe, V. Nickel, B., Giger, T., Müller, U., Groszer, M., Adler, T., Aguilar, A., Bolle, I., Calzada-Wack, J., Dalke, C., Ehrhardt, N., Favor, J., Fuchs, H., Gailus-Durner, V., Hans, W., Hölzlwimmer, G., Javaheri, A., Kalaydjiev, S., Kallnik, M., Kling, E., Kunder, S., Mossbrugger, I., Naton, B., Racz, I., Rathkolb, B., Rozman, J., Schrewe, A., Busch, D. H., Graw, J., Ivandic, B., Klingenspor, M., Klopstock, T., Ollert, M., Quintanilla-Martinez, L., Schulz, H., Wolf, E., Wurst, W., Zimmer, A., Fisher, S. E., Morgenstern, R., Arendt, T., de Angelis, M. H., Fischer, J., Schwarz, J. & Pääbo, S. (2009) A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell* 137:961–71. [SAN]

Engle, R. W. & Kane, M. J. (2004) Executive attention, working memory capacity, and a two-factor theory of cognitive control. *The Psychology of Learning and Motivation* 44:145–99. [KK]

Eraña, A. (2012) Dual process theories versus massive modularity hypotheses. *Philosophical Psychology* 25(6):855–72. [aJMB]

Evans, J. S. B. T. (2011) Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review* 31(2):86–102. [aJMB]

Evans, J. S. B. T. (2013) Two minds rationality. *Thinking & Reasoning* 20:129–46. [aJMB]

Fadiga, L., Craighero, L. & D'Ausilio, A. (2009) Broca's area in language, action, and music. *Annals of the New York Academy of Sciences* 1169(1):448–58. Available at: http://doi.org/10.1111/j.1749-6632.2009.04582.x. [MdJDM]

Fernandes, H. B. F., Woodley, M. A. & te Nijenhuis, J. (2014) Differences in cognitive abilities among primates are concentrated on G: Phenotypic and phylogenetic comparisons with two meta-analytical databases. *Intelligence* 46:311–22. [arJMB, DS]

Fernández-Abascal, E. G., Cabello, R., Fernández-Berrocal, P. & Baron-Cohen, S. (2013) Test-retest reliability of the "Reading the Mind in the Eyes" test: A one-year follow-up study. *Molecular Autism* 4:33. [PF-B]

Fingerman, P. & Levine, M. (1974) Nonlearning: The completeness of the blindness. *Journal of Experimental Psychology* 102:720–21. doi: 10.1037/h0036096. [JOvH]

Fiorito, G. & Scotto, P. (1992) Observational learning in *Octopus vulgaris*. *Science* 256(5056):545. [rJMB]

Firestone, C. & Scholl, B. J. (2015) Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, 1–72. doi: http://dx.doi.org/10.1017/S0140525X15000965. [aJMB]

Fisher, R. A. (1930) *The genetical theory of natural selection.* Oxford University Press. [RA]

Fitch, W. T., Huber, L. & Bugnyar, T. (2010) Social cognition and the evolution of language: Constructing cognitive phylogenies. *Neuron* 65(6):795–814. [LH]

Fitch, W. T. & Martins, M. D. (2014) Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences* 1316 (1):87–104. Available at: http://doi.org/10.1111/nyas.12406. [MdJDM]

Flynn, E. & Whiten, A. (2008) Cultural transmission of tool use in young children. A diffusion chain study. *Social Development* 17:699–718. doi: 10.1111/j.1467-9507.2007.00453.x. [DE]

Flynn, J. R. (2007) *What is intelligence? Beyond the Flynn effect*. Cambridge University Press. Available at: http://books.google.com/books?hl=en&lr=&id=qv-BipuypYUkC&pgis=1. [KK]

Flynn, J. R. (2016) *Does your family make you smarter?: Nature, nurture, and human autonomy*. Cambridge University Press. [rJMB]
Fodor, J. A. (1983) *Modularity of the mind*. MIT Press. [aJMB]
Forss, S. I. F., Schuppli, C., Haiden, D., Zweifel, N. & van Schaik, C. P. (2015) Contrasting responses to novelty by wild and captive orangutans. *American Journal of Primatology* 77(10):1109–21. [arJMB]
Frankenhuis, W. E. & Ploeger, A. (2007) Evolutionary psychology versus Fodor: Arguments for and against the massive modularity hypothesis. *Philosophical Psychology* 20(6):687–710. [aJMB]
Frederick, S. (2005) Cognitive reflection and decision making. *The Journal of Economic Perspectives* 19:25–42. [GP]
Galef, B. G. (2015) Laboratory studies of imitation/field studies of tradition: Towards a synthesis in animal social learning. *Behavioural Processes* 112:114–19. [aJMB]
Galsworthy, M. J., Arden, R. & Chabris, C. F. (2014) Animal models of general cognitive ability for genetic research into cognitive functioning. In: *Behavior genetics of cognition across the lifespan*, ed. D. Finkel & C. A. Reynolds, pp. 257–78. Springer. [aJMB]
Galsworthy, M. J., Paya-Cano, J. L., Liu, L., Monleón, S., Gregoryan, G., Fernandes, C., Schalkwyk, L. C. & Plomin, R. (2005) Assessing reliability, heritability and general cognitive ability in a battery of cognitive tasks for laboratory mice. *Behavioral Genetics* 35(5):675–92. [aJMB]
Galsworthy, M. J., Paya-Cano, J. L., Monleo, S. & Plomin, R. (2002) Evidence for general cognitive ability (g) in heterogeneous stock mice and an analysis of potential confounds. *Genes, Brain and Behavior* 1:88–95. [aJMB]
Garcia, J. & Koelling, R. A. (1966) Relation of cue to consequence in avoidance learning. *Psychonomic Science* 4(1):123–24. [aJMB]
Gärdenfors, P. (2003) *How homo became sapiens: On the evolution of thinking*. Oxford University Press. [IJ]
Gardner, H. (1983/2011) *Frames of mind: The theory of multiple intelligences*. Basic Books. (Original work published in 1983.) [HG]
Gardner, H. (1993) *Multiple intelligences*. Basic Books. [FA]
Gardner, H., Feldman, D. H. & Krechevsky, M. (1998) *Project Spectrum: Preschool Assessment Handbook (Project Zero Frameworks for Early Childhood Education, Volume 3)*. Teachers College Press. [HG]
Garson, D. G. (2013) *Factor analysis*. Statistical Associates. [aJMB]
Geary, D. C. (1995) Reflections of evolution and culture in children's cognition: Implications for mathematical development and instruction. *American Psychologist* 50(1):24. [aJMB]
Geary, D. C. (2005) *The origin of mind. Evolution of brain, cognition, and general Intelligence*. American Psychological Association. [aJMB]
Geary, D. C. (2009) The evolution of general fluid intelligence. In: *Foundations in evolutionary cognitive neuroscience*, ed. S. M. Platek & T. K. Shackelford, pp. 22–56. Cambridge University Press. [aJMB]
Geary, D. C. & Huffman, K. J. (2002) Brain and cognitive evolution: Forms of modularity and functions of mind. *Psychological Bulletin* 128(5):667. [aJMB]
Gelman, R. (1990) First principles organize attention to and learning about relevant data: Number and the animate-inanimate distinction as examples. *Cognitive Science* 14(1):79–106. [aJMB]
Georgas, J., Weiss, L. G., Van de Vijver, F. J. R. & Saklofske, D. H. (2003) *Culture and children's intelligence: Cross-cultural analysis of the WISC-III*. Academic Press. [RJS]
Gergely, G. & Csibra, G. (2006) Sylvia's recipe: The role of imitation and pedagogy in the transmission of human culture. In: *Roots of human sociality: Culture, cognition, and interaction*, ed. N. Enfield & S. Levinson, pp. 229–55. Berg. [DE]
Gervais, W. (2015) Override the controversy: Analytic thinking predicts endorsement of evolution. *Cognition* 142:312–21. [GP]
Gervais, W. M. & Norenzayan, A. (2012) Analytic thinking promotes religious disbelief. *Science* 336:493–96. [GP]
Geva-Sagiv, M., Las, L., Yovel, Y. & Ulanovsky, N. (2015) Spatial cognition in bats and rats: From sensory acquisition to multiscale maps and navigation. *Nature Reviews Neuroscience* 16(2):94–108. Available at: http://doi.org/10.1038/nrn3888. [MdJDM]
Gibson, J. J. (1979) *The ecological approach to visual perception*. Houghton Mifflin. [RJS]
Gilpin, W., Feldman, M. W. & Aoki, K. (2016) An ecocultural model predicts Neanderthal extinction through competition with modern humans. *Proceedings of the National Academy of Sciences USA* 113(8):2134–39. doi: 10.1073/pnas.1524861113. [LDM]
Gläscher, J., Rudrauf, D., Colom, R., Paul, L. K., Tranel, D., Damasio, H. & Adolphs, R. (2010) Distributed neural system for general intelligence revealed by lesion mapping. *Proceedings of the National Academy of Sciences USA* 107 (10):4705–709. [aJMB]
Godfrey-Smith, P. (1996) *Complexity and the function of mind in nature*. Cambridge University Press. [AB]
Godfrey-Smith, P. (2013) Cephalopods and the evolution of the mind. *Pacific Conservation Biology* 19(1): 4–9. [AB]
Goldman, A. & de Vignemont, F. (2009) Is social cognition embodied? *Trends in Cognitive Sciences* 13:154–59. doi: 10.1016/j.tics.2009.01.007. [KEH]
Gottfredson, L. S. (1997) Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence* 24:13–23. [arJMB, CL, FR, LH, MB-H-S]
Graybiel, A. M. (1995) Building action repertoires: Memory and learning functions of the basal ganglia. *Current Opinion in Neurobiology* 5:733–41. [SAN]
Graybiel, A. M. (2008) Habits, rituals, and the evaluative brain. *The Annual Review of Neuroscience* 31:359–87. doi: 10.1146/annurev.neuro.29.051605.112851. [SAN]
Green, M. B. & Schwarz, J. H. (1984) Anomaly cancellations in supersymmetric D = 10 gauge theory and superstring theory. *Physics Letters B* 149(1–3):117–22. Available at: http://doi.org/10.1016/0370-2693(84)91565-X. [RA]
Greenfield, P. M. (1991) Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences* 14:531–95. [MB-H-S]
Greenough, W. T., Black, J. E. & Wallace, C. S. (1987) Experience and brain development. *Child Development* 58:539–59. [aJMB]
Grillner, S., Robertson, B. & Stephenson-Jones, M. (2013) The evolutionary origin of the vertebrate basal ganglia and its role in action selection. *The Journal of Physiology* 591(22):5425–31. [SAN]
Grimm, V. & Mengel, F. (2011) Let me sleep on it: Delay reduces rejection rates in ultimatum games. *Economics Letters* 111:113–15. [GP]
Grossi, G. (2014) A module is a module is a module: Evolution of modularity in Evolutionary Psychology. *Dialectical Anthropology* 38(3):333–51. [aJMB]
Grossman, H. C., Hale, G., Light, K., Kolata, S., Townsend, D. A., Goldfarb, Y., Kusnecov, A. & Matzel, L. D. (2007) Pharmacological modulation of stress reactivity dissociates general learning ability from the propensity for exploration. *Behavioral Neuroscience* 121(5):949. [aJMB]
Gruber, T., Clay, Z. & Zuberbuhler, K. (2010) A comparison of bonobo and chimpanzee tool use: Evidence for a female bias in the Pan lineage. *Animal Behaviour* 80:1023–33. [LH]
Gruber, T., Muller, M. N., Reynolds, V., Wrangham, R. & Zuberbühler, K. (2011) Community-specific evaluation of tool affordances in wild chimpanzees. *Scientific Reports* 1:128. [aJMB]
Guilford, J. P. (1967) *The nature of human intelligence*. McGraw-Hill. [HG]
Gunturkun, O. (2012) The convergent evolution of neural substrates for cognition. *Psychological Research* 76(2):212–19. doi: 10.1007/s00426-011-0377-9. [LDM]
Güntürkün, O. & Bugnyar, T. (2016) Cognition without cortex. *Trends in Cognitive Sciences* 20(4):291–303. [LH]
Gunturkun, O. & Kroner, S. (1999) A polysensory pathway to the forebrain of the pigeon: The ascending projections of the nucleus dorsolateralis posterior thalami (DLP). *European Journal of Morphology* 37(2–3):185–89. [LDM]
Gustafsson, J.-E. (1984) A unifying model for the structure of intellectual abilities. *Intelligence* 8(3):179–203. doi: 10.1016/0160-2896(84)90008-4. [KK]
Gutiérrez-Cobo, M. J., Cabello, R. & Fernández-Berrocal, P. (2016) The relationship between emotional intelligence and cool and hot cognitive processes: A systematic review. *Frontiers in Behavioral Neuroscience* 10:101. [PF-B]
Hagenaars, S. P., Gale, C. R., Deary, I. J. & Harris, S. E. (2016a) Cognitive ability and physical health: A Mendelian randomization study. bioRxiv, 84798. doi: 10.1101/084798. [RCA]
Hagenaars, S. P., Harris, S. E., Davies, G., Hill, W. D., Liewald, D. C. M., Ritchie, S. J., Marioni, R. E., Fawns-Ritchie, C., Cullen, B., Malik, R., METASTROKE Consortium, International Consortium for Blood Pressure GWAS, SpiroMeta Consortium, CHARGE Consortium Pulmonary Group, CHARGE Consortium Aging and Longevity Group, Worrall, B. B., Sudlow, C. L., Wardlaw, J. M., Gallacher, J., Pell, J. McIntosh, A. M., Smith, D. J., Gale, C. R. & Deary, I. J. (2016b) Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112 151) and 24 GWAS consortia. *Molecular Psychiatry* 21:1624–32. doi: 10.1038/mp.2015.225. [RCA]
Haier, R. J., Colom, R., Schroeder, D. H., Condon, C. A., Tang, C., Eaves, E. & Head, K. (2009) Gray matter and intelligence factors: Is there a neuro-g? *Intelligence* 37(2):136–44. doi: 10.1016/j.intell.2008.10.011. [KK]
Halali, E., Bereby-Meyer, Y. & Meiran, N. (2014) Between self-interest and reciprocity: The social bright side of self-control failure. *Journal of Experiment Psychology: General* 143:745–54. [GP]
Hampshire, A., Highfield, R. R., Parkin, B. L. & Owen, A. M. (2012) Fractionating human intelligence. *Neuron* 76:1225–37. [JJL]
Harris, P. L. (2012) *Trusting what you're told: How children learn from others*. Harvard University Press. [DE]
Hauser, M., Pearson, H. & Seelig, D. (2002a) Ontogeny of tool use in cotton-top tamarins, *Saguinus oedipus*: Innate recognition of functionally relevant features. *Animal Behaviour* 64:299–311. [MDH]
Hauser, M. D., Chomsky, N. & Fitch, W. T. (2002b) The faculty of language: What is it, who has it, and how did it evolve? *Science* 298:1569–79. [JJL]
Hauser, M. D. & Watumull, J. (2016) The universal generative faculty: The source of our expressive power in language, mathematics, and music. *Journal of Neurolinguistics*, 1–29. doi: 10.1016/j.jneuroling.2016.10.005. [MDH]
Hayes, S. C., Brownstein, A. J., Zettle, R. D., Rosenfarb, I. & Korn, Z. (1986) Rule-governed behavior and sensitivity to changing consequences of responding.

Journal of the Experimental Analysis of Behavior 45(3):237–56. doi: 10.1901/jeab.1986.45-237. [JOvH]

Healy, S. D. & Rowe, C. (2007) A critique of comparative studies of brain size. *Proceedings of the Royal Society of London B: Biological Sciences* 274 (1609):453–64. [LH]

Heath, A. C., Berg, B., Eaves, L. J., Solass, M. H., Corey, L. A., Sundet, J. M., Magnus, P. & Nance, W. E. (1985) *Nature* 314:734–36. [JJL]

Heldstab, S. A., Kosonen, Z. K., Koski, S. E., Burkart, J. M., van Schaik, C. P. & Isler, K. (2016) Manipulation complexity in primates coevolved with brain size and terrestriality. *Scientific Reports* 6:24528. doi: 10.1038/srep24528. [arJMB, DS]

Henrich, J. (2016) *The secret of our success.* Princeton University Press. [rJMB]

Henrich, J. & Gil-White, F. J. (2001) The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior* 22:165–96. [DMGL]

Herndon, J. G., Moss, M. B., Rosene, D. L. & Killiany, R. (1997) Patterns of cognitive decline in aged rhesus monkeys. *Behavioural Brain Research* 87:25–34. [aJMB]

Herrmann, E. & Call, J. (2012) Are there geniuses among the apes? Philosophical Transactions of the Royal Society *B* 367:2753–61. [aJMB, FA, LH]

Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B. & Tomasello, M. (2007) Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science* 317:1360–66. [arJMB, AW, FA, MAWoM]

Herrmann, E., Hare, B., Call, J. & Tomasello, M. (2010a) Differences in the cognitive skills of bonobos and chimpanzees. *PLoS One* 5(8):e12438. [aJMB, LH]

Herrmann, E., Hernandez-Lloreda, M. V., Call, J., Hare, B. & Tomasello, M. (2010b) The structure of individual differences in the cognitive abilities of children and chimpanzees. *Psychological Science* 21(1):102–10. [aJMB, FA, JOvH, MAWoM]

Heyes, C. (2003) Four routes of cognitive evolution. *Psychological Review* 110 (4):713–27. [aJMB]

Heyes, C. (2012) What's social about social learning? *Journal of Comparative Psychology* 126(2):193–202. [aJMB]

Heyes, C. (2016) Who knows? Metacognitive social learning strategies. *Trends in Cognitive Sciences* 20:204–13. [aJMB]

Heyes, C. M. (1994) Social learning in animals: Categories and mechanisms. *Biological Reviews* 69(2):207–31. [aJMB]

Heyse, L. (2012) Affiliation affects social learning in keas (*Nestor notabilis*) and ravens (*Corvus corax*). (Unpublished master thesis). University of Vienna, Department of Cognitive Biology, Vienna, Austria. [IMP]

Hill, W. D., Davies, G., Harris, S. E., Hagenaars, S. P., The neuroCHARGE Cognitive Working Group, Liewald, D. C., Penke, L., Gale, C. R. & Deary, I. J. (2016a) Molecular genetic aetiology of general cognitive function is enriched in evolutionarily conserved regions. *Translational Psychiatry* 6(12): e980. [RCA]

Hill, B. D., Foster, J. D., Sofko, C., Elliott, E. M. & Shelton, J. T. (2016b) The interaction of ability and motivation: Average working memory is required for Need for Cognition to positively benefit intelligence and the effect increases with ability. *Personality and Individual Differences* 98:225–28. [rJMB]

Hillemann, F., Bugnyar, T., Kotrschal, K. & Wascher, C. A. F. (2014) Waiting for better, not for more: Corvids respond to quality in two delay maintenance tasks. *Animal Behaviour* 90:1–10. [IMP]

Hills, T. T., Todd, P. M. & Goldstone, R. L. (2008) Search in external and internal spaces: Evidence for generalized cognitive search processes. *Psychological Science* 19(8):802–808. [SAN]

Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., Couzin, I. D. & the Cognitive Search Research Group* (*Bateson, M., Cools, R., Dukas, R., Giraldeau, L.-A., Macy, M. W., Page, S. E., Shiffrin, R. M., Stephens, D. W. & Wolfe, J. W.) (2015) Exploration versus exploitation in space, mind, and society. *Trends in Cognitive Sciences* 19(1):46–54. doi: 10.1016/j.tics.2014.10.004. [SAN]

Hochner, B., Shormrat, T. & Fiorito, G. (2006) The octopus: A model for comparative analysis of the evolution of learning and memory mechanisms. *Biological Bulletin* 210:308–17. [KEH]

Hodos, W. & Campbell, C. B. G. (1969) Scala naturae: Why there is no theory in comparative psychology. *Psychological Review* 76(4):337–50. [LH]

Holekamp, K. E. (2007) Questioning the social intelligence hypothesis. *Trends in Cognitive Sciences* 11(2):65–69. [aJMB]

Holekamp, K. E., Dantzer, B., Stricker, G., Yoshida, K. C. S. & Benson-Amram, S. (2015) Brains, brawn and sociality: A hyaena's tale. *Animal Behaviour* 103:237–48. [arJMB]

Hopkins, W. D., Russel, J. L. & Schaeffer, J. (2014) Chimpanzee intelligence is heritable. *Current Biology* 24:1649–52. [aJMB]

Hoppitt, W. & Laland, K. N. (2013) *Social learning: An introduction to mechanisms, methods, and models*. Princeton University Press. [aJMB]

Horn, J. L. & Cattell, R. B. (1967) Age differences in fluid and crystallized intelligence. *Acta Psychologica* 26:107–29. doi: 10.1016/0001-6918(67)90011-X. [KK]

Houle, D. (2000) Is there a g factor for fitness? In: *The nature of intelligence (Novartis Foundation Symposium 233)*, ed. G. Bock, J. Goode & K. Webb, pp. 149–70. Wiley. [RCA]

Howell, R. D., Breivik, E. & Wilcox, J. B. (2007) Reconsidering formative measurement. *Psychological Methods* 12(2):205–18. doi: 10.1037/1082-989X.12.2.205. [KK]

Howrigan, D. P., Simonson, M. A., Davies, G., Harris, S. E., Tenesa, A., Starr, J. M., Liewald, D. C., Deary, I. J., McRae, A., Wright, M. J., Montgomery, G. W., Hansell, N., Martin, N. G., Payton, A., Horan, M., Ollier, W. E., Abdellaoui, A., Boomsma, D. I., DeRosse, P., Knowles, E. E., Glahn, D. C., Djurovic, S., Melle, I., Andreassen, O. A., Christoforou, A., Steen, V. M., Hellard, S. L., Sundet, K., Reinvang, I., Espeseth, T., Lundervold, A. J., Giegling, I., Konte, B., Hartmann, A. M., Rujescu, D., Roussos, P., Giakoumaki, S., Burdick, K. E., Bitsios, P., Donohoe, G., Corley, R. P., Visscher, Pendleton, N., Malhotra, A. K., Neale, B. M., Lencz, T. & Keller, M. C. (2016) Genome-wide autozygosity is associated with lower general cognitive ability. *Molecular Psychiatry* 21(6):837–43. [RA]

Hrdy, S. (2009) *Mothers & others: The evolutionary origins of mutual understanding*. Harvard University Press. [aJMB]

Huber, L. (2009) Degrees of rationality in human and non-human animals. In: *Rational Animals, Irrational Humans*, ed. S. Watanabe, A. P. Blaisdell, L. Huber & A. Young, pp. 3–21. Keio University Press. [LH]

Hufendiek, R. & Wild, M. (2015) Faculties and Modularity. In: *The faculties: A history*, ed. D. Perler, pp. 254–99. Oxford University Press. [aJMB]

Humphrey, N. K. (1976) The social function of intellect. In: *Growing points in ethology*, ed. P. P. G. Bateson & R. A. Hinde, pp. 303–17. Cambridge University Press. [aJMB, IMP]

Hurford, J. R. (2007) *The origins of meaning: Language in the light of evolution*. Oxford University Press. [JJL]

Husby, A. & Husby, M. (2014) Interspecific analysis of vehicle avoidance behavior in birds. *Behavioral Ecology* 25(3):504–508. [aJMB]

Inoue, S. & Matsuzawa, T. (2007) Working memory of numerals in chimpanzees. *Current Biology*. Available at: http://doi.org/10.1016/j.cub.2007.10.027. [MdJDM]

Intelligence and its measurement: A symposium. (1921) *Journal of Educational Psychology* 12:123–47, 195–216, 271–275. [RJS]

Isden, J., Panayi, C., Dingle, C. & Madden, J. (2013) Performance in cognitive and problem-solving tasks in male spotted bowerbirds does not correlate with mating success. *Animal Behaviour* 86(4):829–38. doi: 10.1016/j.anbehav.2013.07.024. [arJMB, JOvH]

Isler, K. & van Schaik, C. P. (2014) How humans evolved large brains: Comparative evidence. *Evolutionary Anthropology: Issues, News, and Reviews* 23(2):65–75. [aJMB]

Iwaniuk, A. N., Lefebvre, L. & Wylie, D. R. (2009) The comparative approach and brain-behavior relationships: A tool for understanding tool use. *Canadian Journal of Experimental Psychology* 63:50–159. [IMP]

Izawa, E. & Watanabe, S. (2008) Formation of linear dominance relationship in captive jungle crows (*Corvus macrorhynchos*): Implications for individual recognition. *Behavioural Processes* 78:44–52. [IMP]

Jackendoff, R. (1999) Possible stages in the evolution of the language capacity. *Trends in Cognitive Science* 3:272–79. [FR]

Jacob, F. (1977) Evolution and tinkering. *Science*, 196:1161–66. doi: 10.1126/science.860134. [RCA]

Jacobs, I. F., von Bayern, A., Martin-Ordas, G., Rat-Fischer, L. & Osvath, M. (2015) Corvids create novel causal interventions after all. *Proceedings of the Royal Society B: Biological Sciences* 282:20142504. doi: 10.1098/rspb.2014.2504. [IJ]

Jaeggi, A. V., Dunkel, L. P., Van Noordwijk, M. A., Wich, S. A., Sura, A. A. & van Schaik, C. P. (2010) Social learning of diet and foraging skills by wild immature Bornean orangutans. *American Journal of Primatology* 72:62–71. [aJMB]

Jarvis, E. D., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., Medina, L., Paxinos, G., Perkel, D., Shimizu, T., Striedter, G., Wild, J. M., Ball, G. F., Dugas-Ford, J., Durand, S. E., Hough, G. E., Husband, S., Kubikova, L., Lee, D. W., Mello, C. V., Powers, A., Siang, C., Smulders, T. V., Wada, K., White, S. A., Yamamoto, K., Yu, J., Reiner, A. & Butler, A. B. (2005) Avian brains and a new understanding of vertebrate evolution. *Nature Reviews Neuroscience* 6:151–59. [IMP]

Jensen, A. R. (1993) Why is reaction time correlated with psychometric g? *Current Directions in Psychological Science* 2:53–56. [FA]

Jensen, A. R. & Weng, L.-J. (1994) What is a good g? *Intelligence* 18(3):231–58. [aJMB]

Jin, X., Tecuapetla, F. & Costa, R. M. (2014) Basal ganglia subcircuits distinctively encode the parsing and concatenation of action sequences. *Nature Neuroscience* 17(3):423–34. doi: 10.1038/nn.3632. [SAN]

Johansen-Berg, H. (2007) Structural plasticity: Rewiring the brain. *Current Biology* 17(4):R141–44. [aJMB]

Johnson, V. E., Deaner, R. O. & van Schaik, C. P. (2002) Bayesian analysis of rank data with application to primate intelligence experiments. *Journal of the American Statistical Association* 97(457):8–17. [aJMB, JJL]

Johnson, W., Penke, L. & Spinath, F. M. (2011) Heritability in the era of molecular genetics: Some thoughts for understanding genetic influences on behavioural traits. *European Journal of Personality* 25:254–66. doi: 10.1002/per.836. [RCA]

Johnson, W., te Nijenhuis, J. & Bouchard, T. J. (2008) Still just 1 g: Consistent results from five test batteries. *Intelligence* 36(1):81–95. [rJMB]

Jolly, A. (1966) Lemur social behaviour and primate intelligence. *Science* 153 (3735):501–506. [aJMB, IMP]

Joshi, P. K., Esko, T., Mattsson, H., Eklund, N., Gandin, I., Nutile, T., Jackson, A. U., Schurmann, C., Smith, A. V., Zhang, W., Okada, Y., Stančáková, A., Faul, J. D., Zhao, W., Bartz, T. M., Concas, M. P., Franceschini, N., Enroth, S., Vitart, V., Trompet, S. & 340 others (2015) Directional dominance on stature and cognition in diverse human populations. *Nature* 523(7561):459–62. [aJMB]

Jung, R. E. & Haier, R. J. (2007) The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behavioral and Brain Sciences* 30(02):135–54. [aJMB]

Kahneman, D. (2011) *Thinking, fast and slow*. Farrar, Straus and Giroux. [GP]

Kahneman, D. & Frederick, S. (2002) Representativeness revisited: Attribute substitution in intuitive judgment. In: *Heuristics and Biases: The Psychology of Intuitive Judgment*, ed. T. Gilovich, D. Griffin & D. Kahneman, pp. 49–81. Cambridge University Press. [FR]

Kahneman, D. & Frederick, S. (2005) A model of heuristic judgement. In: *The Cambridge handbook of thinking and reasoning*, ed. K. J. Holyoak & R. G. Morrison, pp. 267–93. Cambridge University Press. [GP]

Kan, K.-J., Kievit, R. A., Dolan, C. & van der Maas, H. V. (2011) On the interpretation of the CHC factor Gc. *Intelligence* 39(5):292–302. [aJMB]

Kane, M. J. (2005) Full frontal fluidity. In: *Handbook of understanding and measuring intelligence*, ed. O. Wilhelm & R. Engle, pp. 141–65. Sage. [KK]

Kane, M. J. & Engle, R. W. (2002) The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review* 9(4):637–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12613671. [KK]

Kane, M. J., Hambrick, D. Z. & Conway, A. R. A. (2005) Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin* 131:66–71; author reply 72–5. doi: 10.1037/0033-2909.131.1.66. [KK]

Karakuyu, D., Herold, C., Gunturkun, O. & Diekamp, B. (2007) Differential increase of extracellular dopamine and serotonin in the "prefrontal cortex" and striatum of pigeons during working memory. *European Journal of Neuroscience* 26(8):2293–302. [LDM]

Karmiloff-Smith, A. (2015) An alternative to domain-general or domain-specific frameworks for theorizing about human evolution and ontogenesis. *AIMS Neuroscience* 2(2):91–104. Available at: http://doi.org/10.3934/Neuroscience.2015.2.91. [DMGL, MdJDM]

Karten, H. J. (2015) Vertebrate brains and evolutionary connectomics: On the origins of the mammalian "neocortex". *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 370(1684). doi: 10.1098/rstb.2015.0060. [LDM]

Kaufman, A. S. (1979) *Intelligent testing with the WISC-R*. Wiley. [FA]

Kaufman, A. S. (2009) *IQ testing 101*. Springer. [FA]

Kaufman, S. B., DeYoung, C. G., Reis, D. L. & Gray, J. R. (2011) General intelligence predicts reasoning ability even for evolutionarily familiar content. *Intelligence* 39:311–22. [aJMB]

Keagy, J., Savard, J. F. & Borgia, G. (2011) Complex relationship between multiple measures of cognitive ability and male mating success in satin bowerbirds, Ptilonorhynchus violaceus. *Animal Behaviour* 81(5):1063–70. [aJMB, JOvH]

Kell, H. J., Lubinski, D., Benbow, C. P. & Steiger, J. H. (2013) Creativity and technical innovation: Spatial ability's unique role. *Psychological Science* 24:1831–36. [JJL]

Kendal, R. L., Hopper, L. M., Whiten, A., Brosnan, S. F., Lambeth, S. P., Schapiro, S. J. & Hoppitt, W. (2015) Chimpanzees copy dominant and knowledgeable individuals: Implications for cultural diversity. *Evolution and Human Behavior* 36(1):65–72. [aJMB]

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J. & Borsboom, D. (2013) Simpson's Paradox in psychological science: A practical guide. *Frontiers in Psychology* 4: Article 513, 1–14. [aJMB]

Kline, M. A. (2015) How to learn about teaching: An evolutionary framework for the study of teaching behavior in humans and other animals. *Behavioral and Brain Sciences* 38:e31. [aJMB]

Klingberg, T. (2010) Training and plasticity of working memory. *Trends in Cognitive Sciences* 14(7):317–24. [aJMB]

Koenig, A. (1998) Visual scanning by common marmosets (*Callithrix jacchus*): Functional aspects and the special role of adult males. *Primates* 39:85–90. [aJMB]

Koepke, A. E., Gray, S. L. & Pepperberg, I. M. (2015) Delayed gratification: A Grey parrot (*Psittacus erithacus*) will wait for a better reward. *Journal of Comparative Psychology* 129:339–46. [IMP]

Köhler, W. (1925) *The mentality of apes*. Kegan Paul, Trench, Trubner. [JOvH]

Kolata, S., Light, K., Grossman, H. C., Hale, G. & Matzel, L. D. (2007) Selective attention is a primary determinant of the relationship between working memory and general learning ability in outbred mice. *Learning & Memory* 14(1–2):22–28. [aJMB]

Kolata, S., Light, K. & Matzel, L. D. (2008) Domain-specific and domain-general learning factors are expressed in genetically heterogeneous CD-1 mice. *Intelligence* 36(6):619–29. [aJMB, FA]

Kolata, S., Light, K., Townsend, D. A., Hale, G., Grossman, H. C. & Matzel, L. D. (2005) Variations in working memory capacity predict individual differences in general learning abilities among genetically diverse mice. *Neurobiology of Learning and Memory* 84:241–46. [aJMB]

Kolata, S., Light, K., Wass, C. D., Colas-Zelin, D., Roy, D. & Matzel, L. D. (2010) A dopaminergic gene cluster in the prefrontal cortex predicts performance indicative of general intelligence in genetically heterogeneous mice. *PLoS One* 5(11):e14036. [aJMB, LDM]

Kolb, B. & Gibb, R. (2015) Plasticity in the prefrontal cortex of adult rats. *Frontiers in Cellular Neuroscience* 9:15. [aJMB]

Kolodny, O., Edelman, S. & Lotem, A. (2015) Evolved to adapt: A computational approach to animal innovation and creativity. *Current Zoology* 61:350–67. [aJMB]

Kong, D. T. (2014) Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT/MEIS) and overall, verbal, and nonverbal intelligence: Meta-analytic evidence and critical contingencies. *Personality and Individual Differences* 66:171–75. [PF-B]

Koops, K., Furuichi, T. & Hashimoto, C. (2015) Chimpanzees and bonobos differ in intrinsic motivation for tool use. *Scientific Reports* 5:11356. [aJMB]

Korman, J., Voiklis, J. & Malle, B. F. (2015) The social life of cognition. *Cognition* 135:30–35. [aJMB]

Kovacs, K. & Conway, A. R. A. (2016) Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry* 27(3):151–77. doi: 10.1080/1047840X.2016.1153946. [KK, rJMB]

Kovacs, K., Plaisted, K. C. & Mackintosh, N. J. (2006) Difficulties differentiating dissociations. *Behavioral and Brain Sciences* 29(02):138–39. doi: 10.1017/S0140525X06349035. [KK]

Kovas, Y. & Plomin, R. (2006) Generalist genes: Implications for the cognitive sciences. *Trends in Cognitive Sciences* 10:198–203. [JJL]

Krakauer, E. B. (2005) *Development of Aye-Aye* (Daubentonia madagascariensis) foraging skills: Independent exploration and social learning. Duke University. [aJMB]

Krijnen, W. P. (2004) Positive loadings and factor correlations from positive covariance matrices. *Psychometrika* 69(4):655–60. doi: 10.1007/BF02289861. [KK]

Kumaran, D., Melo, H. L. & Düzel, E. (2012) The emergence and representation of knowledge about social and nonsocial hierarchies. *Neuron* 76(3):653–66. Available at: http://doi.org/10.1016/j.neuron.2012.09.035.1. [MdJDM]

Kuzawa, C. W., Chugani, H. T., Grossman, L. I., Lipovich, L., Muzik, O., Hof, P. R., Wildman, D. E., Sherwood, C. C., Leonard, W. R. & Lange, N. (2014) Metabolic costs and evolutionary implications of human brain development. *Proceedings of the National Academy of Sciences USA* 111(36):13010–15. [aJMB]

Laland, K. N., Sterelny, K., Odling-Smee, J., Hoppitt, W. & Uller, T. (2011) Cause and effect in biology revisited: Is Mayr's proximate-ultimate dichotomy still useful? *Science* 334:1512–16. doi: 10.1126/science.1210879. [IJ]

Lea, S. E. G. & Wills, A. J. (2008) Use of multiple dimensions in learned discriminations. *Comparative Cognition & Behavior Reviews* 3:115–33. doi: 10.3819/ccbr.2008.30007. [JOvH]

Lea, S. E. G., Wills, A. J., Leaver, L. A., Ryan, C. M. E., Bryant, C. M. L. & Millar, L. (2009) A comparative analysis of the categorization of multidimensional stimuli: II. Strategic information search in humans (Homo sapiens) but not in pigeons (Columba livia). *Journal of Comparative Psychology* 123(4):406–20. doi: 10.1037/a0016851. [JOvH]

Lee, J. (2007) A g beyond Homo sapiens? Some hints and suggestions. *Intelligence* 35:253–65. [FA]

Lee, J. J. (2012) Correlation and causation in the study of personality (with discussion). *European Journal of Personality* 26:372–412. [JJL]

Lee, J. J. & Chow, C. C. (2013) The causal meaning of Fisher's average effect. *Genetics Research* 95:89–109. [JJL]

Lee, J. J. & Kuncel, N. R. (2015) The determinacy and predictive power of common factors. *Industrial and Organizational Psychology* 8:467–72. [JJL]

Lee, J. J., Vattikuti, S. & Chow, C. C. (2016) Uncovering the genetic architectures of quantitative traits. *Computational and Structural Biotechnology Journal* 14:28–34. [JJL]

Lefebvre, L. (2013) Brains, innovations, tools and cultural transmission in birds, non-human primates, and fossil hominins. *Frontiers in Human Neuroscience* 7:245. [aJMB]

Lefebvre, L. (2014) Should neuroecologists separate Tinbergen's four questions? *Behavioural Processes* 117:92–96. [aJMB]

Lefebvre, L., Reader, S. M. & Sol, D. (2004) Brains, innovations and evolution in birds and primates. *Brain, Behavior and Evolution* 63:233–46. [arJMB]

Lefebvre, L., Reader, S. M. & Sol, D. (2013) Innovating innovation rate and its relationship with brains, ecology and general intelligence. *Brain, Behavior and Evolution* 81:143–45. [aJMB, FA, KEH]

Lefebvre, L., Whittle, P., Lascaris, E. & Finkelstein, A. (1997) Feeding innovations and forebrain size in birds. *Animal Behaviour* 53:549–60. [aJMB]

Legare, C. H. & Harris, P. L. (2016) Introduction to special section on the ontogeny of cultural learning. *Child Development* 87(3):633–42. [DE]

Legree, P. J., Psotka, J., Roberts, R. D., Robbins, J., Putka, D. J. & Mullins, H. M. (2014) Profile similarity metrics as an alternate framework to score rating-based tests: MSCEIT reanalyses. *Intelligence* 47:159–74. [PF-B]

Lesage, E., Navarrete, G. & De Neys, W. (2013) Evolutionary modules and Bayesian facilitation: The role of general cognitive resources. *Thinking & Reasoning* 19 (1):27–53. [aJMB]

Lewis, D. M. G., Al-Shawaf, L., Conroy-Beam, D., Asao, K. & Buss, D. M. (2017) Evolutionary psychology: A how-to guide. *American Psychologist* 72(4):353–73. [DMGL]

Light, K. R., Grossman, Y., Kolata, S., Wass, C. D. & Matzel, L. D. (2011) General learning ability regulates exploration through its influence on rate of habituation. *Behavioral Brain Research* 223:297–309. [aJMB, LDM]

Light, K. R., Kolata, S., Hale, G., Grossman, H. & Matzel, L. D. (2008) Up-regulation of exploratory tendencies does not enhance general learning abilities in juvenile or young-adult outbred mice. *Neurobiology of Learning and Memory* 90:317–29. [aJMB]

Light, K. R., Kolata, S., Wass, C., Denman-Brice, A., Zagalsky, R. & Matzel, L. D. (2010) Working memory training promotes general cognitive abilities in genetically heterogeneous mice. *Current Biology* 20:777–82. [aJMB]

Locurto, C. (1991) *Sense and nonsense about IQ: The case for uniqueness*. Praeger Press. [CL]

Locurto, C. (1997) On the comparative generality of *g*. In: *Advances in cognition and education, vol. 4: Reflections on the concept of intelligence*, ed. W. Tomic & J. Kigman, pp. 79–100. JAI Press. [aJMB, CL]

Locurto, C. & Scanlon, C. (1998) Individual differences and a spatial learning factor in two strains of mice (*Mus musculus*). *Journal of Comparative Psychology* 112 (4):344–52. [aJMB, CL]

Locurto, C., Benoit, A., Crowley, C. & Miele, A. R. (2006) The structure of individual differences in batteries of rapid acquisition tasks in mice. *Journal of Comparative Psychology* 120:378–88. [aJMB, CL, LH]

Locurto, C., Fortin, E. & Sullivan, R. (2003) The structure of individual differences in heterogeneous stock mice across problem types and motivational systems. *Genes, Brain and Behavior* 2(1):40–55. [aJMB]

Loehlin, J. C., Hansell, N. K., Wright, M. J. & Martin, N. G. (2016) Genetic and environmental contributions to cognitive structure in Australian twins: A reappraisal. *Intelligence* 59:24–31. [JJL]

Logan, C. J. (2016a) Behavioral flexibility and problem solving in an invasive bird. *PeerJ* 4:1–30, e1975. doi: 10.7717/peerj.1975. [AB]

Logan, C. J. (2016b) How far will a behaviourally flexible invasive bird go to innovate? *Royal Society Open Science* 3(6):1–10. [AB]

Logan, C. J., Jelbert, S. A., Breen, A. J., Gray, R. D. & Taylor, A. H. (2014) Modifications to the Aesop's fable paradigm change New Caledonian crow performances. *PLoS One* 9(7):1–9, e103049. doi: 10.1371/journal.pone.0103049. [AB]

Lotem, A. & Halpern, J. Y. (2012) Coevolution of learning and data-acquisition mechanisms: A model for cognitive evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 367(1603):2686–94. [aJMB]

Lovejoy, A. O. (1936) *The great chain of being: A study of the history of an idea*. Harvard University Press. [RJS]

Luncz, L. V. & Boesch, C. (2014) Tradition over trend: Neighboring chimpanzee communities maintain differences in cultural behavior despite frequent immigration of adult females. *American Journal of Primatology* 76(7):649–57. [aJMB]

Luria, A. R. (1976) *Cognitive development: Its cultural and social foundations*. Harvard University Press. [RJS]

MacCann, C., Joseph, D. L., Newman, D. A. & Roberts, R. D. (2014) Emotional intelligence is a second-stratum factor of intelligence: Evidence from hierarchical and bifactor models. *Emotion* 14:358–74. [PF-B]

Mackintosh, N. J. (1998) *IQ and human intelligence*. Oxford University Press. [CL]

Mackintosh, N. J. (2011) *IQ and human intelligence* (2nd ed.). Oxford University Press. [FR]

MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., Aureli, F., Baker, J. M., Bania, A. E., Barnard, A. M., Boogert, N. J., Brannon, E. M., Bray, E. E., Bray, J., Brent, L. J. N., Burkart, J. M., Call, J., Cantlon, J. F., Cheke, L. G., Clayton, N. S., Delgado, M. M., DiVinventi, L. J., Fujita, K., Herrmann, E., Hiramatsu, C., Jacobs, L. F., Jordan, K. E., Laude, J. R., Leimgruber, K. L., Messer, E. J. E., Moura, A. C. de, A., Ostojic, L., Picard, A., Platt, M. L., Plotnik, J. M., Range, F., Reader, S. M., Reddy, R. B., Sandel, A. A., Santos, L. R., Schumann, K., Seed, A. M., Sewall, K. B., Shaw, R. C., Slocombe, K. E., Yanjie, S., Takimoto, A., Tan, J., Tao, R., van Schaik, C. P., Viranyi, Z., Visalberghi, E., Wade, J. C., Watanabe, A., Widness, J., Young, J. K., Zentall, T. R. & Zhao, Y. (2014) The evolution of self-control. *Proceedings of the National Academy of Sciences USA* 111:E2140–48. [aJMB, GP]

Macphail, E. M. (1982) *Brain and intelligence in vertebrates*. Clarendon Press. [aJMB, CL]

Macphail, E. M. (1987) The comparative psychology of intelligence. *Behavioral and Brain Sciences* 10:645–56. [CL]

Maes, E., De Filippo, G., Inkster, A. B., Lea, S. E. G., De Houwer, J., D'Hooge, R., Beckers, T. & Wills, A. J. (2015) Feature- versus rule-based generalization in rats, pigeons and humans. *Animal Cognition* 18(6):1267–84. doi: 10.1007/s10071-015-0895-8. [JOvH]

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. & Frith, C. D. (2000) Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences USA* 97(8):4398–403. [LDM]

Mahon, B. Z. & Cantlon, J. F. (2011) The specialization of function: Cognitive and neural perspectives. *Cognitive Neuropsychology* 28(3–4):147–55. [aJMB]

Major, J. T., Johnson, W. & Deary, I. J. (2012) Comparing models of intelligence in Project TALENT: The VPR model fits better than the CHC and extended Gf–Gc models. *Intelligence* 40(6):543–59. [aJMB]

Malle, B. F. (2002) The relation between language and theory of mind in development and evolution. In: *The evolution of language out of pre-language*, ed. T. Givon & B. F. Malle, pp. 265–84. Benjamins. [DE]

Martinho, A. & Kacelnik, A. (2016) Ducklings imprint on the relational concept of "same or different". *Science* 353:286–88. doi: 10.1126/science.aaf4247. [IJ, MDH]

Martins, M. D., Fischmeister, F. P., Puig-Waldmüller, E., Oh, J., Geißler, A., Robinson, S., Fitch, W. T. & Beisteiner, R. (2014) Fractal image perception provides novel insights into hierarchical cognition. *NeuroImage* 96:300–08. Available at: http://doi.org/10.1016/j.neuroimage.2014.03.064. [MdJDM]

Mather, J. A. (1994) "Home" choice and modification by juvenile *Octopus vulgaris*: Specialized intelligence and tool use? *Journal of Zoology* 233:359–68. [KEH]

Mather, J. A. & Anderson, R. C. (1999) Exploration, play and habituation in *Octopus dofleini*. *Journal of Comparative Psychology* 113:333–38. [KEH]

Mather, J. A. & Kuba, M. J. (2013) The cephalopod specialties: Complex nervous system, learning, and cognition. *Canadian Journal of Zoology* 91:431–49. [AB]

Mather, J. A., Leite, T. & Battista, A. T. (2012) Individual prey choices of octopus: Are they generalists or specialists? *Current Zoology* 58:597–603. [KEH]

Matsunaga, E., Nambu, S., Oka, M., Tanaka, M., Taoka, M. & Iriki, A. (2015) Identification of tool use acquisition-associated genes in the primate neocortex. *Development, Growth & Differentiation* 57(6):484–95. [aJMB]

Matzel, L. D., Grossman, H., Light, K., Townsend, D. & Kolata, S. (2008) Age-related declines in general cognitive abilities of Balb/C mice are associated with disparities in working memory, body weight, and general activity. *Learning & Memory* 15(10):733–46. [aJMB]

Matzel, L. D., Han, Y. R., Grossmann, H., Karnik, M. S., Patel, D., Scott, N., Specht, S. M. & Gandhi, C. C. (2003) Individual differences in the expression of a "general" learning ability in mice. *The Journal of Neuroscience* 23(16):6423–33. [aJMB]

Matzel, L. D., Light, K. R., Wass, C., Colas-Zelin, D., Denman-Brice, A., Waddel, A. C. & Kolata, S. (2011a) Longitudinal attentional engagement rescues mice from age-related cognitive declines and cognitive inflexibility. *Learning and Memory* 18(5):345–56. [aJMB]

Matzel, L. D., Sauce-Silva, B. & Wass, C. (2013) The architecture of intelligence: Converging evidence from studies of humans and animals. *Current Directions in Psychological Science* 22:342–48. [aJMB, LDM]

Matzel, L. D., Townsend, D. A., Grossman, H., Han, Y. R., Hale, G., Zappulla, M., Light, K. & Kolata, S. (2006) Exploration in outbred mice covaries with general learning abilities irrespective of stress reactivity, emotionality, and physical attributes. *Neurobiology of Learning and Memory* 86:228–40. [aJMB, LDM]

Matzel, L. D., Wass, C. & Kolata, S. (2011b) Individual differences in animal intelligence: Learning, reasoning, selective attention and inter-species conservation of a cognitive trait. *International Journal of Comparative Psychology* 24:36–59. [aJMB, CL]

Matzke, D., Dolan, C. V. & Molenaar, D. (2010) The issue of power in the identification of "g" with lower-order factors. *Intelligence* 38(3):336–44. doi: 10.1016/j.intell.2010.02.001. [KK]

Mayer, J. D., Roberts, R. D. & Barsade, S. G. (2008) Human abilities: Emotional intelligence. *Annual Review of Psychology* 59:507–36. [PF-B]

Mayer, J. D. & Salovey, P. (1997) What is emotional intelligence? In: *Emotional development and emotional intelligence: Implications for educators*, ed. P. Salovey & D. Sluyter, pp. 3–31. Basic Books. [PF-B]

Mayer, J. D., Salovey, P. & Caruso, D. R. (2002) *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) User's Manual*. Multi-Health Systems. [PF-B]

McDaniel, M. A. (2005) Big-brained people are smarter: A meta-analysis of the relationship between in vivo brain volume and intelligence. *Intelligence* 33:337–46. [LDM]

McGrew, K. S. (2009) CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence* 37(1):1–10. doi: 10.1016/j.intell.2008.08.004. [aJMB, KK, PF-B]

McNab, F., Varrone, A., Farde, L., Jucaite, A., Bystritsky, P., Forssberg, H. & Klingberg, T. (2009) Changes in cortical dopamine D1 receptor binding associated with cognitive training. *Science* 323(5915):800–802. [LDM]

Meier, C., Lea, S. E. G. & McLaren, I. P. L. (2016) A stimulus-location effect in contingency-governed, but not rule-based, discrimination learning. *Journal of Experimental Psychology: Animal Learning and Cognition* 42(2):177–86. doi: 10.1037/xan0000098. [JOvH]

Mellars, P. & French, J. C. (2011) Tenfold population increase in Western Europe at the Neandertal–to–modern human transition. *Science* 333(6042):623–27. [rJMB]

Melnick, M. D., Harrison, B. R., Park, S., Bennetto, L. & Tadin, D. (2013) A strong interactive link between sensory discriminations and intelligence. *Current Biology* 23(11):1013–17. [aJMB]

Meltzoff, A. N. (1995) What infant memory tells us about infantile amnesia: Long-term recall and deferred imitation. *Journal of Experimental Child Psychology* 59:497–515. [DE]

Meltzoff, A. N., Kuhl, P. K., Movellan, J. & Sejnowski, T. J. (2009) Foundations for a new science of learning. *Science* 325(5938):284–88. [DE]

Mendes, N. & Huber, L. (2004) Object permanence in common marmosets (Callithrix jacchus). *Journal of Comparative Psychology* 118(1):103–12. [LH]

Meulman, E. J. M., Sanz, C. M., Visalberghi, E. & van Schaik, C. P. (2012) The role of terrestriality in promoting primate technology. *Evolutionary Anthropology: Issues, News, and Reviews* 21(2):58–68. [aJMB]

Meulman, E. J. M., Seed, A. M. & Mann, J. (2013) If at first you don't succeed … Studies of ontogeny shed light on the cognitive demands of habitual tool use. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 368(1630):20130050 [aJMB]

Miller, E. K. & Cohen, J. D. (2001) An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience* 24:167–202. doi: 10.1146/annurev.neuro.24.1.167. [LDM]

Miller, R., Schiestl, M., Whiten, A., Schwab, C. & Bugnyar, T. (2014) Tolerance and social facilitation in the foraging behaviour of free-ranging crows (*Corvus corone corone*; *C. c. cornix*). *Ethology* 120:1248–55. [IMP]

Mills, C. M. (2013) Knowing when to doubt: Developing a critical stance when learning from others. *Developmental Psychology* 49(3). doi: 10.1037/a0029500. [DE]

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A. & Wager, T. D. (2000) The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology* 41(1):49–100. [aJMB]

Moll, H. & Tomasello, M. (2007) Cooperation and human cognition: The Vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 362(1480):639–48. [aJMB, AW]

Montgomery, S. (2015) *The Soul of an Octopus*. Simon & Schuster. [KEH]

Morand-Ferron, J. & Quinn, J. L. (2015) The evolution of cognition in natural populations. *Trends in Cognitive Sciences* 19(5):235–37. [rJMB]

Morand-Ferron, J., Cole, E. F. & Quinn, J. L. (2015) Studying the evolutionary ecology of cognition in the wild: A review of practical and conceptual challenges. *Biological Reviews* 91(2):367–89. [rJMB]

Morrison, A. B. & Chein, J. M. (2011) Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review* 18(1):46–60. [aJMB]

Muthukrishna, M. & Henrich, J. (2016) Innovation in the collective brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*. [aJMB]

Nippak, P. M. & Milgram, N. W. (2005) An investigation of the relationship between response latency across several cognitive tasks in the beagle dog. *Progress in Neuro-psychopharmacology & Biological Psychiatry* 29(3):371–77. [aJMB]

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F. & Turkheimer, E. (2012) Intelligence: New findings and theoretical developments. *American Psychologist* 67(2):130–59. [arJMB]

Noack, H., Lövdén, M. & Schmiedek, F. (2014) On the validity and generality of transfer effects in cognitive training research. *Psychological Research* 78:773–89. doi: 10.1007/s00426-014-0564-6. [RCA]

Nordli, S. A. (2012) The evolutionary economics of learning: The lawful structure of habit formation and the computational capacity for recursion in behavior (Unpublished undergraduate thesis). School of Cognitive Science, Hampshire College, Amherst, MA. Available from Harold F. Johnson Library, Hampshire College. [SAN]

O'Hara, M., Auersperg, A. M. I., Bugnyar, T. & Huber, L. (2015) Inference by exclusion in Goffin cockatoos (*Cacatua goffini*). *PLoS One* 10(8):e0134894. [LH]

O'Hara, M., Schwing, R., Federspiel, I., Gajdon, G. K. & Huber, L. (2016) Reasoning by exclusion in the kea (*Nestor notabilis*). *Animal Cognition* 19(5):965–75. [LH]

O'Neill, D. K. (1996) Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development* 67:659–77. [DE]

Oberauer, K., Schulze, R., Wilhelm, O. & Süss, H.-M. (2005) Working memory and intelligence—Their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin* 131:61–5; author reply 72–5. doi: 10.1037/0033-2909.131.1.61. [KK, rJMB]

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., Turley, P., Chen, G-B., Emilsson, V., Meddens, S. Fleur, W., Oskarsson, S., Pickrell, J. K., Thom, K., Timshel, P., de Vlaming, R., Abdellaoui, A., Ahluwalia, T. S., Bacelis, J., Baumbach, C., Bjornsdottir, G., Brandsma, J. H., Pin Concas, M., Derringer, J., Furlotte, N. A., Galesloot, T. E., Girotto, G., Gupta, R., Hall, L. M., Harris, S. E., Hofer, E., Horikoshi, M., Huffman, J. E., Kaasik, K., Kalafati, I. P., Karlsson, R., Kong, A., Lahti, J., van der Lee, S. J., deLeeuw, C., Lind, P. A., Liu, T., Mangino, M., Marten, J., Mihailov, E., Miller, M. B., van der Most, P. J. Oldmeadow, C., Payton, A., Pervjakova, N., Peyrot, W. J. Qian, Y., Raitakari, O., Rueedi, R., Salvi, E., Schmidt, B., Schraut, K. E., Shi, J., Smith, A. V., Poot, R. A., St Pourcain, B., Teumer, A., Thorleifsson, G., Verweij, N., Vuckovic, D., Wellmann, J., Westra, H.-J., Yang, J., Zhao, W., Zhu, Z., Alizadeh, B. Z., Amin, N., Bakshi, A., Baumeister, S. E., Biino, G., Bønnelykke, K., Boyle, P. A., Campbell, H., Cappuccio, F. P., Davies, G., De Neve, J.-E., Deloukas, P., Demuth, I., Ding, J., Eibich, P., Eisele, L., Eklund, N., Evans, D. M., Faul, J. D., Feitosa, M. F., Forstner, A. J., Gandin, I., Gunnarsson, B., Halldórsson, B. V., Harris, T. B., Heath, A. C., Hocking, L. J., Holliday, E. G., Homuth, G., Horan, M. A., Hottenga, J.-J., de Jager, P. L., Joshi, P. K., Jugessur, A., Kaakinen, M. A., Kähönen, M., Kanoni, S., Keltigangas-Järvinen, L., Kiemeney, L. A. L. M., Kolcic, I., Koskinen, S., Kraja, A. T., Kroh, M., Kutalik, Z., Latvala, A., Launer, L. J., Lebreton, M. P., Levinson, D. F., Lichtenstein, P., Lichtner, P., Liewald, D. C. M. LifeLines Cohort Study, Loukola, A., Madden, P. A., Mäki-Opas, T., Marioni, R. E., Marques-Vidal, P., Meddens, G. A., McMahon, G., Meisinger, C., Meitinger, T., Milaneschi, Y., Milani, L., Montgomery, G. W., Myhre, R., Nelson, C. P., Nyholt, D. R., Ollier, W. E. R., Palotie, A., Paternoster, L., Pedersen, N. L., Petrovic, K. E., Porteous, D. J., Räkkönen, K., Ring, S. M., Robino, A., Rostapshova, O., Rudan, I., Rustichini, A., Salomaa, V., Sanders, A. R., Sarin, A.-P., Schmidt, H., Scott, R. J., Smith, B. H., Smith, J. A., Staessen, J. A., Steinhagen-Thiessen, E., Strauch, K., Terracciano, A., Tobin, M. D., Ulivi, S., Vaccargiu, S., Quaye, L., van Rooij, F. J. A., Venturini, C., Vinkhuyzen, A. A. E., Völker, U., Völzke, H., Vonk, J. M., Vozzi, D., Waage, J., Ware, E. B., Willemsen, G., Attia, J. R., Bennett, D. A., Berger, K., Bertram, L., Bisgaard, H., Boomsma, D. I., Boreck, I. B., Bültmann, U., Chabris, C. F., Cucca, F., Cusi, D., Deary, I. J., Dedoussis, G. V., van Duijn, C. M., Eriksson, J. G., Franke, B., Franke, L., Gasparini, P., Gejman, P. V., Gieger, C., Grabe, H.-J., Gratten, J., Groenen, P. J. F., Gudnason, V., van der Harst, P., Hayward, C., Hinds, D. A., Hoffmann, W., Hyppönen, E., Iacono, W. G., Jacobsson, B., Järvelin, M.-R., Jöckel, K.-H., Kaprio, J., Kardia, S. L. R., Lehtimäki, T., Lehrer, S. F., Magnusson, P. K. E., Martin, N. G., McGue, M., Metspalu, A., Pendleton, N., Penninx, B. W. J. H., Perola, M., Pirastu, N., Pirastu, M., Polasek, O., Posthuma, D., Power, C., Province, M. A., Samani, N. J., Schlessinger, D., Schmidt, R., Sørensen, T. I. A., Spector, T. D., Stefansson, K., Thorsteinsdottir, U., Thurik, A. R., Timpson, N. J., Tiemeier, H., Tung, J. Y., Uitterlinden, A. G., Vitart, V., Vollenweider, P., Weir, D. R., Wilson, J. F., Wright, A. F., Conley, D. C., Krueger, R. F., Smith, G. D., Hofman, A., Laibson, D. I., Medland, S. E., Meyer, M. N., Yang, J., Johannesson, M., Visscher, P. M., Esko, T., Koillinger, P. D., Cesarini, D. & Benjamin, D. B. (2016) Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 533:539–42. [JJL]

Olkowicz, S., Kocourek, M., Lučan, R. K., Porteš, M., Fitch, W. T., Herculano-Houzel, S. & Němec, P. (2016) Birds have primate-like numbers of neurons in the forebrain. *Proceedings of the National Academy of Sciences USA* 113 (26):7255–60. doi: 10.1073/pnas.1517131113. [IMP, rJMB]

Ortiz, S. O. (2015) CHC theory of intelligence. In: *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts*, ed. S. Goldstein, D. Princiotta & J.A. Naglieri, pp. 209–27. Springer. [aJMB]

Osvath, M., Kabadayi, C. & Jacobs, I. F. (2014) Independent evolution of similar complex cognitive skills: The importance of embodied degrees of freedom. *Animal Behavior and Cognition* 1:249–64. doi: 10.12966/abc.08.03.2014. [IJ]

Oudeyer, P. Y., Kaplan, F. & Hafner, V. (2007) Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation* 11:265–86. doi: 10.1109/TEVC.2006.890271. [KEH]

Owen, K. (1992) The suitability of Raven's Standard Progressive Matrices for various groups in South Africa. *Personality and Individual Differences* 13(2):149–59. [MB-H-S]

Paap, K. R., Johnson, H. A. & Sawi, O. (2015) Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex* 69:265–78. doi: 10.1016/j.cortex.2015.04.014. [RCA]

Pahl, M., Si, A. & Zhang, S. (2013) Numerical cognition in bees and other insects. *Frontiers in Psychology* 4:162. doi: 10.3389/fpsyg.2013.00162. [MB-H-S]

Parker, S. T. (2015) Re-evaluating the extractive foraging hypothesis. *New Ideas in Psychology* 37:1–12. [aJMB]

Parker, S. T. & Gibson, K. R. (1977) Object manipulation, tool use and sensorimotor intelligence as feeding adaptations in *Cebus* monkeys and great apes. *Journal of Human Evolution* 6(7):623–41. [aJMB, MB-H-S]

Patton, B. W. & Braithwaite, V. A. (2015) Changing tides: Ecological and historical perspectives on fish cognition. *Wiley Interdisciplinary Reviews: Cognitive Science* 6(2):159–76. [aJMB]

Pavlicev, M. & Wagner, G. P. (2012) Coming to grips with evolvability. *Evolution: Education and Outreach* 5(2):231–44. [aJMB]

Paz-y-Miño, G., Bond, A. B. & Kamil, A. C. (2004) Pinyon jays use transitive inference to predict social dominance. *Nature* 430:778–81. [IMP]

Penke, L., Borsboom, D., Johnson, W., Kievit, R. A., Ploeger, A. & Wicherts, J. M. (2011) Evolutionary psychology and intelligence research cannot be integrated the way Kanazawa (2010) suggested. *American Psychologist* 66:916–17. doi: 10.1037/a0024626. [RCA]

Penke, L., Denissen, J. J. A. & Miller, G. F. (2007) The evolutionary genetics of personality. *European Journal of Personality* 21:549–87. doi: 10.1002/per.629. [RCA]

Pennington, B. F., Filipek, P. A., Lefly, D., Chhabildas, N., Kennedy, D. N., Simon, J. H., Filley, C. M., Galaburda, A. & DeFries, J. C. (2000) A twin MRI study of size variations in the human brain. *Journal of Cognitive Neuroscience* 12:223–32. [FA]

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J. & Fugelsang, J. A. (2014) The role of analytic thinking in moral judgements and values. *Thinking & Reasoning* 20:188–214. [GP]

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J. & Fugelsang, J. A. (2015a) On the reception and detection of pseudo-profound bullshit. *Judgment and Decision Making* 10:549–63. [GP]

Pennycook, G., Cheyne, J. A., Koehler, D. J. & Fugelsang, J. A. (2016a) Is the cognitive reflection test a measure of both reflection and intuition? *Behavior Research Methods* 48:341–48. [GP]

Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J. & Fugelsang, J. A. (2012) Analytic cognitive style predicts religious and paranormal belief. *Cognition* 123:335–46. [GP]

Pennycook, G., Fugelsang, J. A. & Koehler, D. J. (2015b) Everyday consequences of analytic thinking. *Current Directions in Psychological Science* 24:425–43. [GP]

Pennycook, G., Ross, R., Koehler, D. J. & Fugelsang, J. A. (2016b) Atheists and agnostics are more reflective than religious believers: Four empirical studies and a meta-analysis. *PLoS One* 11:e0153039. [GP]

Pepperberg, I. M. (1981) Functional vocalizations by an African Grey parrot. *Zeitschrift für Tierpsychologie* 55:139–60. [IMP]

Pepperberg, I. M. (1983) Cognition in the African Grey parrot: Preliminary evidence for auditory/vocal comprehension of the class concept. *Animal Learning & Behavior* 11:179–85. [IMP]

Pepperberg, I. M. (1987) Acquisition of the same/different concept by an African Grey parrot (*Psittacus erithacus*): Learning with respect to color, shape, and material. *Animal Learning & Behavior* 15:423–32. [IMP]

Pepperberg, I. M. (1988) Comprehension of "absence" by an African Grey parrot: Learning with respect to questions of same/different. *Journal of the Experimental Analysis of Behavior* 50:553–64. [IMP]

Pepperberg, I. M. (1999) *The Alex studies*. Harvard University Press. [IMP]

Pepperberg, I. M. (2006) Ordinality and inferential abilities of a Grey parrot (*Psittacus erithacus*). *Journal of Comparative Psychology* 120:205–16. [IMP]

Pepperberg, I. M. & Brezinsky, M. V. (1991) Acquisition of a relative class concept by an African Grey parrot (*Psittacus erithacus*): Discriminations based on relative size. *Journal of Comparative Psychology* 105:286–94. [IMP]

Pepperberg, I. M. & Carey, S. (2012) Grey parrot number acquisition: The inference of cardinal value from ordinal position on the numeral list. *Cognition* 125:219–32. [IMP]

Pepperberg, I. M. & Gordon, J. D. (2005) Number comprehension by a Grey parrot (*Psittacus erithacus*), including a zero-like concept. *Journal of Comparative Psychology* 119:197–209. [IMP]

Pepperberg, I. M., Koepke, A., Livingston, P., Girard, M. & Hartsfield, L. A. (2013) Reasoning by inference: Further studies on exclusion in Grey parrots (*Psittacus erithacus*). *Journal of Comparative Psychology* 127:272–81. [IMP]

Phelps, E. A. & LeDoux, J. E. (2005) Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron* 48(2):175–87. [X-TW]

Phelps, E. A., Lempert, K. M. & Sokol-Hessner, P. (2014) Emotion and decision making: Multiple modulatory neural circuits. *Annual Review of Neuroscience* 37:263–87. [PF-B]

Piaget, J. (1952) *The origins of intelligence in children*. International Universities Press. [MB-H-S]

Piaget, J. (1971) *Biology and knowledge*. University of Chicago Press. [MB-H-S]

Pietschnig, J., Penke, L., Wicherts, J. M., Zeiler, M. & Voracek, M. (2015) Meta-analysis of associations between human brain volume and intelligence differences: How strong are they and what do they mean? *Neuroscience and Biobehavioral Reviews* 57:411–32. [aJMB]

Pinker, S. (1997) *How the mind works*. W. W. Norton. [MDH]

Pinker, S. (2010) The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences USA* 107(Supplement 2):8993–99. [aJMB]

Pinker, S. (2014) Why academics stink at writing. *Chronicle of Higher Education* 61 (5):1–16. Available at: http://stevenpinker.com/files/pinker/files/why_academics_stink_at_writing.pdf. [RA]

Pinker, S. & Bloom, P. (1990) Natural language and natural selection. *Behavioral and Brain Sciences* 13:707–84. [FR]

Pinker, S. & Jackendoff, R. (2005) The faculty of language: What's special about it? *Cognition* 95:201–36. [JJL]

Plakke, B., Hwang, J. & Romanski, L. M. (2015) Inactivation of primate prefrontal cortex impairs auditory and audiovisual working memory. *Journal of Neuroscience* 35(26):9666–75. Available at: http://doi.org/10.1523/JNEUROSCI.1218-15.2015. [MdJDM]

Ploeger, A. & Galis, F. (2011) Evo devo and cognitive science. *Wiley Interdisciplinary Reviews: Cognitive Science* 2:429–40. doi: 10.1002/wcs.137. [aJMB, IJ]

Plomin, R. (2001) The genetics of g in human and mouse. *Nature Reviews Neuroscience* 2(2):136–41. [aJMB]

Pravosudov, V. V. & Roth II, T. C. (2013) Cognitive ecology of food hoarding: The evolution of spatial memory and the hippocampus. *Annual Review of Ecology, Evolution, and Systematics* 44:173–93. [aJMB]

Prinz, J. J. (2006) Is the mind really modular? In: *Contemporary Debates in Cognitive Science*, ed. R. J. Stainton, pp. 22–36. Blackwell. [aJMB]

Prokosch, M. D., Yeo, R. A. & Miller, G. F. (2005) Intelligence tests with higher g loadings show higher correlations with body symmetry: Evidence for a general fitness factor mediated by developmental stability. *Intelligence* 33:203–13. [DMGL]

Quartz, S. (2003) Toward a developmental evolutionary psychology: Genes, development and the evolution of cognitive architecture. In: *Evolutionary psychology: Alternative approaches*, ed. S. J. Scher & F. Rauscher, pp. 185–210. Kluwer. [aJMB]

Rabipour, S. & Raz, A. (2012) Training the brain: Fact and fad in cognitive and behavioral remediation. *Brain and Cognition* 79(2):159–79. [aJMB]

Raby, C. R., Alexis, D. M., Dickinson, A. & Clayton, N. S. (2007) Planning for the future by western scrub-jays. *Nature* 445:919–21. [IMP]

Rand, D. G. (2016) Cooperation, fast and slow: Meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychological Science* 27:1192–206. doi: 10.1177/0956797616654455. [GP]

Rand, D. G., Brescoll, V. L., Everett, J. A. C., Capraro, V. & Barcelo, H. (2016) Social heuristics and social roles: Altruism is intuitive for women but not for men. *Journal of Experiment Psychology: General* 145:389–96. [GP]

Rand, D. G., Greene, J. D. & Nowak, M. A. (2012) Spontaneous giving and calculated greed. *Nature* 498:427–30. [GP]

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A. & Greene, J. D. (2014) Social heuristics shape intuitive cooperation. *Nature Communications* 5:3677. doi: 10.1038/ncomms4677. [GP]

Reader, S. M. (2003) Innovation and social learning: Individual variation and brain evolution. *Animal Biology* 53:147–58. [aJMB]

Reader, S. M., Hager, Y. & Laland, K. N. (2011) The evolution of primate general and cultural intelligence. *Philosophical Transactions of the Royal Society B* 366:1017–27. [arJMB, DS]

Reader, S. M. & Laland, K. N. (2002) Social intelligence, innovation and enhanced brain size in primates. *Proceedings of the National Academy of Sciences USA*, 99:4436–41. [aJMB, LH]

Redick, T. S. (2015) Working memory training and interpreting interactions in intelligence interventions. *Intelligence* 50:14–20. doi: 10.1016/j.intell.2015.01.014. [RCA]

Reeve, C. L. (2004) Differential ability antecedents of general and specific dimensions of declarative knowledge: More than *g*. *Intelligence* 32(6):621–52. [aJMB]

Reiner, A. (2010) The conservative evolution of the vertebrate basal ganglia. In: *Handbook of basal ganglia structure and function*, ed. H. Steiner & K. Y. Tseng, pp. 29–62. Academic Press. [SAN]

Reyes-García, V., Pyhälä, A., Díaz-Reviriego, I., Duda, R., Fernández-Llamazares, Á., Gallois, S., Guèze, M. & Napitupulu, L. (2016) Schooling, local knowledge and working memory: A study among three contemporary hunter-gatherer societies. *PloS One* 11(1):e0145265. [aJMB]

Rhodes, G., Nishimura, M., de Heering, A., Jeffery, L. & Maurer, D. (2017) Reduced adaptability, but no fundamental disruption, of norm-based face coding following early visual deprivation from congenital cataracts. *Developmental Science* 20:e12384. doi: 10.1111/desc.12384. [MDH]

Richerson, P. J., Baldini, R., Bell, A., Demps, K., Frost, K., Hillis, V., Mathew, S., Newton, N., Narr, N., Newson, L., Ross, C., Smaldino, P., Waring, T. & Zefferman, M. (2016) Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. *Behavioral and Brain Sciences* 39:e30. [aJMB]

Riddell, W. I. & Corl, K. G. (1977) Comparative investigation of the relationship between cerebral indices and learning abilities. *Brain, Behavior and Evolution* 14:385–98. [aJMB]

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X. & Behrens, T. E. J. (2008) The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience* 11(4):426–28. Available at: http://doi.org/10.1038/nn2072. [MdJDM]

Robbins, M. M., Ando, C., Fawcett, K. A., Gruiter, C. C., Hedwig, D., Iwata, Y., Lodwick, J. L., Masi, S., Salmi, R., Stoinski, T. S., Todd, A., Vercellio, V. & Yamagiwa, J. (2016) Behavioural variation in gorillas: Evidence of potential cultural traits. *PLoS One* 11:e0160483. [AW]

Rosati, A. G. & Santos, L. R. (2016) Spontaneous metacognition in Rhesus monkeys. *Psychological Science* 27:1181–91. doi: 10.1177/0956797616653737. [GP]
Rosenzweig, M. R. & Bennett, E. L. (1996) Psychobiology of plasticity: Effects of training and experience on brain and behavior. *Behavioral Brain Research* 78 (1):57–65. [LDM]
Roth, G. (2013) *The long evolution of brains and minds*. Dordrecht. [AB]
Rowe, C. & Healy, S. D. (2014) Measuring variation in cognition. *Behavioral Ecology* 25(6):1287–92. [rJMB]
Royall, D. R. & Palmer, R. F. (2014) "Executive functions" cannot be distinguished from general intelligence: Two variations on a single theme within a symphony of latent variance. *Frontiers in Behavioral Neuroscience* 8(369):1–10. [aJMB]
Royzman, E. B., Landy, J. F. & Goodwin, G. P. (2014) Are good reasoners more incest-friendly? Trait cognitive reflection predicts selective moralization in a sample of American adults. *Judgment and Decision Making* 9:176–90. [GP]
Rozin, P. (1976) The evolution of intelligence and access to the cognitive unconscious. In: *Progress in psychobiology and physiological psychology*, ed. J. M. Sprague & A. N. Epstein, pp. 245–80. Academic Press. [MB-H-S]
Rumbaugh, D. M. & Washburn, D. A. (2003) *Intelligence of apes and other rational beings*. Yale University Press. [aJMB]
Saklofske, D. H., van de Vijver, F. J., Oakland, T., Mpofu, E. & Suzuki, L. A. (2014) Intelligence and culture: History and assessment. In: *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts*, ed. S. Goldstein, D. Princiotta & J.A. Naglieri, pp. 341–65. Springer. [aJMB]
Sale, A., Berardi, N. & Maffei, L. (2014) Environment and brain plasticity: Towards an endogenous pharmacotherapy. *Physiological Reviews* 94(1):189–234. [aJMB]
Samuels, R. (2004) Innateness in cognitive science. *Trends in Cognitive Sciences* 8 (3):136141. [aJMB]
Santorelli, C. J., Schaffner, C. M., Campbell, C. J., Notman, H., Pavelka, M. S., Weghorst, J. A. & Aureli, F. (2011) Traditions in wild spider monkeys are biased towards the social domain. *PLoS One* 6:e16863. [AW]
Sauce, B., Wass, C., Smith, A., Kwan, S. & Matzel, L. D. (2014) The external-internal loop of interference: Two types of attention and their influence on the learning abilities of mice. *Neurobiology of Learning and Memory* 116:181–92. [aJMB, CL]
Savage-Rumbaugh, S., Fields, W. M., Segerdahl, P. & Rumbaugh, D. (2005) Culture prefigures cognition in Pan/Homo bonobos. *Theoria. Revista de Teoría, Historia y Fundamentos de la Ciencia* 20(3):311–28. [aJMB]
Schloegl, C. (2011) What you see is what you get – Reloaded: Can jackdaws (*Corvus monedula*) find hidden food through exclusion? *Journal of Comparative Psychology* 125:162–74. [IMP]
Schloegl, C., Dierks, A., Gajdon, G. K., Huber, L., Kotrschal, K. & Bugnyar, T. (2009) What you see is what you get? Exclusion performances in ravens and keas. *PLoS One* 4:e6368. [IMP]
Schlosser, G. & Wagner, G. P. (2004) *Modularity in development and evolution*. University of Chicago Press. [aJMB]
Schmidhuber, J. A. (1991) Possibility for implementing curiosity and boredom in model-building neural controllers. In: *Proceedings of the International Conference on Simulation of Adaptive Behavior: From Animals to Animats*, pp. 222–27. [KEH]
Schmitt, V., Pankau, B. & Fischer, J. (2012) Old world monkeys compare to apes in the primate cognition test battery. *PLoS One* 7(4):e32024. [FA]
Schreiweis, C., Bornschein, U., Burguière, E., Kerimoglu, C., Schreiter, S., Dannemann, M., Goyal, S., Rea, E., French, C. A., Puliyadi, R., Groszer, M., Fisher, S. E., Mundry, R., Winter, C., Hevers, W., Pääbo, S., Enard, W. & Graybiel, A. M. (2014) Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *Proceedings of the National Academy of Sciences USA* 111(39):14253–58. doi: 10.1073/pnas.1414542111. [SAN]
Schubiger, M. N., Kissling, A. & Burkart, J. B. (2016) How task format affects cognitive performance: A memory test with two species of New World monkeys. *Animal Behaviour* 121:33–39. [aJMB]
Schubiger, M. N., Wüstholz, F. L., Wunder, A. & Burkart, J. M. (2015) High emotional reactivity toward an experimenter affects participation, but not performance, in cognitive tests with common marmosets (Callithrix jacchus). *Animal Cognition* 18(3):701–12. [aJMB]
Schuppli, C., Isler, K. & van Schaik, C. P. (2012) How to explain the unusually late age at skill competence among humans. *Journal of Human Evolution* 63(6):843–50. [aJMB]
Schuppli, C., Meulman, E., Forss, S. F., van Noordwijk, M. & van Schaik, C. P. (2016) Observational social learning and socially induced practice of routine skills in wild immature orang-utans. *Animal Behaviour* 119:87–98. [aJMB]
Scott, B. H., Mishkin, M. & Yin, P. B. (2012) Monkeys have a limited form of short-term memory in audition. *Proceedings of the National Academy of Sciences USA* 109(30):12237–41. Available at: http://doi.org/10.1073/pnas.1209685109. [MdJDM]
Seashore, C. E. (1923) *Introduction to psychology*. Macmillan. [LDM]
Seligman, M. E. (1970) On the generality of the laws of learning. *Psychological Review* 77(5):406–18. doi: 10.1037/h0029790. [JOvH]
Seyfarth, R. M. & Cheney, D. (2014) The evolution of language from social cognition. *Current Opinion in Neurobiology* 28:5–9. Available at: http://doi.org/10.1016/j.conb.2014.04.003. [MdJDM]
Shah, P., Happé, F., Sowden, S., Cook, R. & Bird, G. (2015) Orienting toward face-like Stimuli in early childhood. *Child Development* 86(6):1693–700. [aJMB]
Sharkey, N. E. & Ziemke, T. (1998) A consideration of the biological and psychological foundations of autonomous robotics. *Connection Science* 10:361–91. doi: 10.1080/095400998116495. [KEH]
Shaw, R. C., Boogert, N. J., Clayton, N. S. & Burns, K. C. (2015) Wild psychometrics: Evidence for "general" cognitive performance in wild New Zealand robins, Petroica longipes. *Animal Behaviour* 109:101–11. doi: 10.1016/j.anbehav.2015.08.001. [arJMB, JOvH]
Shenhav, A., Rand, D. G. & Greene, J. D. (2012) Divine intuition: Cognitive style influences belief in God. *Journal of Experimental Psychology: General* 141:423–28. [GP]
Shepherd, S. V. (2010) Following gaze: Gaze-following behavior as a window into social cognition. *Frontiers in Integrative Neuroscience* 4:5. [aJMB]
Sheppard, L. D. & Vernon, P. A. (2008) Intelligence and speed of information-processing: A review of 50 years of research. *Personality and Individual Differences* 44(3):535–51. [aJMB]
Sherry, D. F. (2006) Neuroecology. *Annual Review of Psychology* 57:167–97. [aJMB]
Shettleworth, S. J. (2010a) Clever animals and killjoy explanations in comparative psychology. *Trends in Cognitive Sciences* 14(11):477–81. [LH]
Shettleworth, S. J. (2010b) *Cognition, evolution, and behavior*. Oxford University Press. [FA]
Shettleworth, S. J. (2012a) Darwin, Tinbergen, and the evolution of comparative cognition. In: *The Oxford handbook of comparative evolutionary psychology*, ed. J. Vonk & T. K. Shackelford, pp. 529–46. Oxford University Press. [aJMB]
Shettleworth, S. J. (2012b) Modularity, comparative cognition and human uniqueness. *Philosophical Transactions of the Royal Society B* 367(1603):2794–802. [aJMB]
Shipstead, Z., Redick, T. & Engle, R. W. (2012) Is working memory training effective? *Psychological Bulletin* 138(4):628. [aJMB]
Shtulman, A. & McCallum, K. (2014) Cognitive reflection predicts science understanding. In: *Proceedings of the 36th Annual Conference of the Cognitive Science Society, Quebec, ON, July 2014*, pp. 2937–42. Erlbaum. [GP]
Shultz, S. & Dunbar, R. I. M. (2006) Both social and ecological factors predict ungulate brain size. *Proceedings of the Royal Society B: Biological Sciences* 273 (1583):207–15. [aJMB]
Shumway, C. A. (2008) Habitat complexity, brain, and behavior. *Brain, Behavior and Evolution* 72(2):123–34. [aJMB]
Silberman, D. (2016) *Neural tribes*. Avery/Penguin. [HG]
Sinn, D. L., Perrin, N. A., Mather, J. A. & Anderson, R. C. (2001) Early temperamental traits in an octopus (*Octopus bimaculoides*). *Journal of Comparative Psychology* 115:351–64. [KEH]
Smirnova, A. A. (2013) Symbolic representation of the numerosities 1-8 by hooded crows (*Corvus cornix L.*). Paper presented at the International Ethological Congress, Newcastle, UK, August 4–8. [IMP]
Smirnova, A., Zorina, Z., Obozova, T. & Wasserman, E. (2015) Crows spontaneously exhibit analogical reasoning. *Current Biology* 25:256–60. [MDH]
Smith, J. D., Ashby, F. G., Berg, M. E., Murphy, M. S., Spiering, B., Cook, R. G. & Grace, R. C. (2011) Pigeons' categorization may be exclusively nonanalytic. *Psychonomic Bulletin & Review* 18(2):414–21. doi: 10.3758/s13423-010-0047-8. [JOvH]
Smith, J. D., Berg, M. E., Cook, R. G., Murphy, M. S., Crossley, M. J., Boomer, J., Spiering, B., Beran, M. J., Church, B. A., Ashby, F. G. & Grace, R. C. (2012) Implicit and explicit categorization: A tale of four species. *Neuroscience and Biobehavioral Reviews* 36(10):2355–69. doi: 10.1016/j.neubiorev.2012.09.003. [JOvH]
Smith, K. S. & Graybiel, A. M. (2016) Habit formation. *Dialogues in Clinical Neuroscience* 18(1):33–43. [SAN]
Smith, P. K. (1996) Language and the evolution of mind-reading. In: *Theories of theories of mind*, ed. P. Carruthers & P. K. Smith, pp. 344–54. Cambridge University Press. [DE]
Sol, D. (2009a) Revisiting the cognitive buffer hypothesis for the evolution of large brains. *Biology Letters* 5:130–33. [FA, KEH]
Sol, D. (2009b) The cognitive-buffer hypothesis for the evolution of large brains. In: *Cognitive. Ecology II*, ed. R. Dukas & J. M. Ratcliffe pp. 111–34. University of Chicago Press. [aJMB, KEH]
Sol, D., Bacher, S., Reader, S. M. & Lefebvre, L. (2008) Brain size predicts the success of mammal species introduced into novel environments. *The American Naturalist* 172(S1):S63–71. [aJMB]
Sol, D., Duncan, R. P., Blackburn, T. M., Cassey, P. & Lefebvre, L. (2005) Big brains, enhanced cognition, and response of birds to novel environments. *Proceedings of the National Academy of Sciences USA* 102(15):5460–65. [aJMB]
Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M. & Smoller, J. W. (2013) Pleiotropy in complex traits: Challenges and strategies. *Nature Reviews Genetics* 14:483–95. doi: 10.1038/nrg3461. [RCA]

Soltész, F., Goswami, U., White, S. & Szűcs, D. (2011) Executive function effects and numerical development in children: Behavioural and ERP evidence from a numerical Stroop paradigm. *Learning and Individual Differences* 21:662–71. [MDH]

Spearman, C. (1927) *The abilities of man*. Macmillan. xxii 415pp. [aJMB]

Spearman, C. (1904) "General intelligence," objectively determined and measured. The American Journal of Psychology 15:201–92. [CL, FR]

Spearman, C. (1927) *The abilities of man*. Macmillan. [CL, FA]

Spelke, E. S. & Kinzler, K. D. (2007) Core knowledge. *Developmental Science* 10 (1):89–96. [aJMB]

Sperber, D. (2001) In defense of massive modularity. In: *Language, brain and cognitive development: Essays in honor of Jacques Mehler*, ed. E. Dupoux, pp. 47–57. MIT Press. [aJMB]

Spitz, H. H. (1988) Wechsler subtest patterns of mentally retarded groups: Relationship to *g* and to estimates of heritability. *Intelligence* 12:279–97. [MAWoM]

Stanovich, K. E. (2009a) Is it time for a tri-process theory? Distinguishing the reflective and algorithmic mind. In: *In two minds: Dual processes and beyond*, ed. J. St. B. T. Evans & K. Frankish, pp. 55–88. Oxford University Press. [GP]

Stanovich, K. E. (2009b) *What intelligence tests miss: The psychology of rational thought*. Yale University Press. [GP]

Stanovich, K. E. (2011) *Rationality and the reflective mind*. Oxford University Press. [GP]

Stanovich, K. E. & West, R. F. (1998) Individual differences in rational thought. *Journal of Experimental Psychology: General* 127:161–88. [GP]

Stanovich, K. E. & West, R. F. (2000) Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences* 23:645–726. [GP]

Stanton, C. & Clune, J. (2016) Curiosity search: Producing generalists by encouraging individuals to continually explore and acquire skills throughout their lifetime. *PLoS One* 11(9):e0162235. doi: 10.1371/journal.pone.0162235. [KEH]

Stephens, D. W. & Krebs, J. R. (1986) *Foraging theory*. Princeton University Press. [SAN]

Stephenson-Jones, M., Samuelsson, E., Ericsson, J., Robertson, B. & Grillner, S. (2011) Evolutionary conservation of the basal ganglia as a common vertebrate mechanism for action selection. *Current Biology* 21:1081–91. doi: 10.1016/j.cub.2011.05.001. [SAN]

Sterelny, K. (2003) *Thought in a hostile world: The evolution of human cognition*. Blackwell. [AB]

Sternberg, R. J. (1984) *Beyond IQ*. Cambridge University Press. [HG]

Sternberg, R. J. (1988) *The triarchic mind: A new theory of human intelligence*. Viking. [FA]

Sternberg, R. J. (1990) *Metaphors of mind: Conceptions of the nature of intelligence*. Cambridge University Press. [RJS]

Sternberg, R. J. (2004) Culture and intelligence. *American Psychologist* 59(5):325–38. [RJS]

Sternberg, R. J. (ed.) (2002) *Why smart people can be so stupid*. Yale University Press. [RJS]

Sternberg, R. J. & Detterman, D. K. (ed.) (1986) *What is intelligence?* Ablex. [RJS]

Sternberg, R. J., Grigorenko, E. L. & Kidd, K. K. (2005) Intelligence, race, and genetics. *American Psychologist* 60(1):46–59. [RJS]

Sternberg, S. (2011) Modular processes in mind and brain. *Cognitive Neuropsychology* 28(3–4):156–208. [aJMB]

Stevens, J. P. (2012) *Applied multivariate statistics for the social sciences*. Routledge. [aJMB]

Stevens, J. R., Kennedy, B. A., Morales, D. & Burks, M. (2016) The domain specificity of intertemporal choice in pinyon jays. *Psychonomic Bulletin & Review* 23(3):915–21. [rJMB]

Stocco, A., Lebiere, C. & Anderson, J. R. (2010) Conditional routing of information to the cortex: A model of the basal ganglia's role in cognitive coordination. *Psychological Review* 117(2):541–74. doi: 10.1037/a0019077. [SAN]

Strasser, A. & Burkart, J. M. (2012) Can we measure brain efficiency? An empirical test with common marmosets (Callithrix jacchus). *Brain, Behavior and Evolution* 80(1):26–40. [aJMB]

Sutter, M., Kocher, M. & Strauß, S. (2003) Bargaining under time pressure in an experimental ultimatum game. *Economics Letters* 81:341–47. [GP]

Swami, V., Voracek, M., Stieger, S., Tran, U. S. & Furnham, A. (2014) Analytic thinking reduces belief in conspiracy theories. *Cognition* 133:572–85. [GP]

Szabo, B., Bugnyar, T. & Auersperg, A. M. I. (2016) Within-group relationships and lack of social enhancement during object manipulation in captive Goffin's cockatoos (*Cacatua goffiniana*). *Learning & Behavior* 45(1):7–19. doi: 10.3758/s13420-016-0235-0. [IMP]

Tapp, P. D., Siwak, C. T., Estrada, J., Head, E., Muggenburg, B. A., Cotman, C. W. & Milgram, N. W. (2003) Size and reversal learning in the beagle dog as a measure of executive function and inhibitory control in aging. *Learning & Memory* 10(1):64–73. [aJMB]

ten Cate, C. (2016) Assessing the uniqueness of language: Animal grammatical abilities take center stage. *Psychonomic Bulletin & Review* 1–6. doi: 10.3758/s13423-016-1091-9. [MDH]

Teschke, I., Cartmill, E. A., Stankewitz, S. & Tebbich, S. (2011) Sometimes tool use is not the key: No evidence for cognitive adaptive specializations in tool-using woodpecker finches. *Animal Behaviour* 82(5):945–56. [LH]

Teschke, I., Wascher, C. A., Scriba, M. F., von Bayern, A. M., Huml, V., Siemers, B. & Tebbich, S. (2013) Did tool-use evolve with enhanced physical cognitive abilities? *Philosophical Transactions of the Royal Society B: Biological Sciences* 368(1630):20120418. [LH]

Teubert, M., Vierhaus, M. & Lohaus, A. (2011) Frühkindliche Untersuchungsmethoden zur Intelligenzprognostik. *Psychologische Rundschau* 62(2):70–77. [aJMB]

Thomson, G. H. (1916) A hierarchy without a general factor. *British Journal of Psychology* 8:271–81. [FR]

Thornton, A. & Lukas, D. (2012) Individual variation in cognitive performance: Developmental and evolutionary perspectives. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 367:2773–83. doi: 10.1098/rstb.2012.0214. [RCA]

Thorsen, C., Gustafsson, J. E. & Cliffordson, C. (2014) The influence of fluid and crystallized intelligence on the development of knowledge and skills. *British Journal of Educational Psychology* 84(4):556–70. [aJMB]

Tinbergen, N. (1963) On aims and methods of ethology. *Zeitschrift für Tierpsychologie* 20(4):410–33. [aJMB]

Toates, F. (2005) Evolutionary psychology–towards a more integrative model. *Biology and Philosophy* 20(2–3):305–28. [aJMB]

Tomasello, M. (1999) *The cultural origins of human cognition*. Harvard University Press. [aJMB]

Tomasello, M. (2016) Cultural learning redux. *Child Development* 87(3):643–53. [DE]

Tomasello, M. & Call, J. (1997) *Primate cognition*. Oxford University Press. [arJMB, IJ]

Tomlin, D. A., Rand, D. G., Ludvig, E. A. & Cohen, J. D. (2015) The evolution and devolution of cognitive control: The costs of deliberation in a competitive world. *Scientific Reports* 5:11002. doi: 10.1038/srep11002. [GP]

Tooby, J. & Cosmides, L. (1990) On the universality of human nature and the uniqueness of the individual: The role of genetics and adaptation. Journal of Personality 58:17–67. [MAWoM]

Toupo, D. F. P., Strogatz, S. H., Cohen, J. D. & Rand, D. G. (2015) Evolutionary game dynamics of controlled and automatic decision-making. *Chaos* 25:073120. Available at: http://dx.doi.org/10.1063/1.4927488. [GP]

Trahan, L. H., Stuebing, K. K., Fletcher, J. M. & Hiscock, M. (2014) The Flynn effect: A meta-analysis. *Psychological Bulletin* 140:1332–60. [KK]

Trzaskowski, M., Davis, O. S. P., DeFries, J. C., Yang, J., Visscher, P. M. & Plomin, R. (2013) DNA evidence for strong genome-wide pleiotropy of cognitive and learning abilities. *Behavior Genetics* 43:267–73. [JJL]

Ujfalussy, D., Miklosi, A., Bugnyar, T. & Kotrschal, K. (2014) Role of mental representations in quantity judgements by jackdaws (*Corvus monedula*). *Journal of Comparative Psychology* 128:11–20. [IMP]

Vallortigara, G. (2012a) Core knowledge of object, number, and geometry: A comparative and neural approach. *Cognitive Neuropsychology* 29:213–36. doi: 10.1080/02643294.2012.654772. [IJ]

Vallortigara, G. (2012b) The cognitive chicken: Visual and spatial cognition in a non–mammalian brain. In: *The Oxford handbook of comparative cognition*, ed. T. R. Zentall & E. A. Wasserman, pp. 48–66. Oxford University Press. [IJ]

van de Waal, E., Borgeaud, C. & Whiten, A. (2013) Potent social learning and conformity shape a wild primate's foraging decisions. *Science* 340(6131):483–85. [aJMB]

van der Maas, H. L. J., Dolan, C. V, Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M. & Raijmakers, M. E. J. (2006) A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review* 113 (4):842–61. doi: 10.1037/0033-295X.113.4.842. [arJMB, DS, FR, KK]

van der Sluis, S., Verhage, M., Posthuma, D. & Dolan, C. (2010) Phenotypic complexity, measurement bias, and poor phenotypic resolution contribute to the missing heritability problem in genetic association studies. *PLoS One* 5:e13929. [JJL]

van Horik, J. O. & Emery, N. J. (2016) Transfer of physical understanding in a non-tool-using parrot. *Animal Cognition* 19:1195–1203. [IMP]

van Leeuwen, M., Peper, J. S. & van den Berg, S. M. (2009) A genetic analysis of brain volumes and IQ in children. *Intelligence* 37:417–24. [LDM]

van Praag, H., Kempermann, G. & Gage, F. H. (2000) Neural consequences of environmental enrichment. *Nature Reviews Neuroscience* 1(3):191–98. [LDM]

van Schaik, C. P. (2016) *The primate origins of human nature*. Wiley-Blackwell. [aJMB]

van Schaik, C. P., Ancrenaz, M., Borgen, G., Galdikas, B., Knott, C. D., Singleton, I., Suzuki, A., Suci, S. U. & Merrill, M. (2003) Orangutan cultures and the evolution of material culture. *Science* 299:102–105. [aJMB, AW]

van Schaik, C. P. & Burkart, J. M. (2011) Social learning and evolution: The cultural intelligence hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366(1567):1008–16. [aJMB, AW, FA]

van Schaik, C. P., Burkart, J. M., Damerius, L., Forss, S. I. F., Koops, K., van Noordwijk, M. A. & Schuppli, C. (2016) The reluctant innovator: Orangutans

and the phylogeny of creativity. *Philosophical Transactions B* 371 (1690):20150183. [arJMB]

van Schaik, C. P., Deaner, R. O. & Merrill, M. Y. (1999) The conditions for tool use in primates: Implications for the evolution of material culture. *Journal of Human Evolution* 36:719–41. [aJMB]

van Schaik, C. P., Graber, S. M., Schuppli, C. & Burkart, J. M. (2017) The ecology of social learning in animals and its link with intelligence. *Spanish Journal of Comparative Psychology*. Published online. doi: https://doi.org/10.1017/sjp.2016.100. [aJMB]

van Schaik, C. P., Isler, K. & Burkart, J. M. (2012) Explaining brain size variation: From social to cultural brain. *Trends in Cognitive Sciences* 16:277–84. [aJMB, FA, KEH, LDM]

Van Valen, L. (1974) Brain size and intelligence in man. *American Journal of Physical Anthropology* 40:417–24. [LDM]

van Woerden, J. T., van Schaik, C. P. & Isler, K. (2010) Effects of seasonality on brain size evolution: Evidence from strepsirrhine primates. *The American Naturalist* 176(6):758–67. [aJMB]

van Woerden, J. T., van Schaik, C. P. & Isler, K. (2014) Brief communication: Seasonality of diet composition is related to brain size in New World Monkeys. *American Journal of Physical Anthropology* 154(4):628–32. [aJMB]

van Woerden, J. T., Willems, E. P., van Schaik, C. P. & Isler, K. (2012) Large brains buffer energetic effects of seasonal habitats in catarrhine primates. *Evolution* 66 (1):191–99. [arJMB]

Veit, L., Hartmann, K. & Nieder, A. (2014) Neuronal correlates of visual working memory in the corvid endbrain. *Journal of Neuroscience* 34(23):7778–86. doi: 10.1523/JNEUROSCI.0612-14.2014. [LDM]

Vicari, S., Bellucci, S. & Carlesimo, G. A. (2007) Visual and spatial working memory dissociation: Evidence from Williams syndrome. *Developmental Medicine & Child Neurology* 45(4):269–73. doi: 10.1111/j.1469-8749.2003.tb00342.x. [KK]

Voelkl, B. & Huber, L. (2000) True imitation in marmosets. *Animal Behaviour* 60 (2):195–202. [LH]

Voelkl, B. & Huber, L. (2007) Imitation as faithful copying of a novel technique in marmoset monkeys. *PLoS One* 2(7):e611. [LH]

Voelkl, B., Schrauf, C. & Huber, L. (2006) Social contact influences the response of infant marmosets towards novel food. *Animal Behaviour* 72:365–72. [aJMB]

Wang, P. P. & Bellugi, U. (1994) Evidence from two genetic syndromes for a dissociation between verbal and visual-spatial short-term memory. *Journal of Clinical and Experimental Neuropsychology* 16(2):317–22. doi: 10.1080/01688639408402641. [KK]

Wang, X. T. (1996) Domain-specific rationality in human choices: Violations of utility axioms and social contexts. *Cognition* 60(1):31–63. [X-TW]

Wass, C., Denman-Brice, A., Light, K. R., Kolata, S., Smith, A. M. & Matzel, L. D. (2012) Covariation of learning and "reasoning" abilities in mice: Evolutionary conservation of the operations of intelligence. *Journal of Experimental Psychology: Animal Behavior Processes* 38(2):109–24. [aJMB, LH]

Wass, C., Pizzo, A., Sauce, B., Kawasumi, Y., Sturzoiu, T., Ree, F., Otto, T. & Matzel, L. D. (2013) Dopamine D1 sensitivity in the prefrontal cortex predicts general cognitive abilities and is modulated by working memory training. *Learning & Memory* 20(11):617–27. [LDM]

Webb, C. A., Schwab, Z. J., Weber, M., DelDonno, S., Kipman, M., Weiner, M. R. & Killgore, W. D. (2013) Convergent and divergent validity of integrative versus mixed model measures of emotional intelligence. *Intelligence* 41:149–56. [PF-B]

Wechsler, D., Coalson, D. L. & Raiford, S. E. (2008) *WAIS-IV: Wechsler adult intelligence scale*. Pearson. [LH]

Wellman, H. M. (1990) *The child's theory of mind.* MIT Press. [DE]

Wellman, H. M., Cross, D. & Watson, J. (2001) Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development* 72:655–84. [DE]

West-Eberhard, M. J. (2003) *Developmental plasticity and evolution*. Oxford University Press. [IJ]

Whitehead, H. & Rendell, L. (2015) *The cultural lives of whales and dolphins*. Chicago University Press. [AW]

Whiten, A. (2015) Experimental studies illuminate the cultural transmission of percussive technology in Homo and Pan. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370:20140359. [AW]

Whiten, A. (2017) Social learning and culture in child and chimpanzee. *Annual Review of Psychology* 68:129–64. doi: 10.1146/annurev-psych-010416-044108. [AW]

Whiten, A. (in press) The evolution and ontogeny of "deep social mind" and the social brain. In: *Minnesota Symposia on Child Psychology, Vol. 39, Development of the Social Brain*, ed. J. T. Elison & M. D. Sera. Wiley. [AW]

Whiten, A. & van Schaik, C. P. (2007) The evolution of animal "cultures" and social intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences* 362(1480):603–20. [aJMB, AW]

Whiten, A., Goodall, J., McGrew, W.C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C. E. G., Wrangham, R. W. & Boesch, C. (1999) Cultures in chimpanzees. *Nature* 399:682–85. [AW]

Whiten, A., McGuigan, N., Marshall-Pescini, S. & Hopper, L. M. (2009) Emulation, imitation, overimitation and the scope of culture for child and chimpanzee. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364:2417–28. [AW]

Wickett, J. C., Verbnon, P. A. & Lee, D. H. (2000) Relationships between factors of intelligence and brain volume. *Personality and Individual Differences* 29:1095–122. [LDM]

Will, B., Galani, R., Kelche, C. & Rosenzweig, M. R. (2004) Recovery from brain injury in animals: Relative efficacy of environmental enrichment, physical exercise or formal training (1990–2002). *Progress in Neurobiology* 72(3):167–82. doi: 10.1016/j.pneurobio.2004.03.001. [LDM]

Wills, A. J., Lea, S. E. G., Leaver, L. A, Osthaus, B., Ryan, C. M. E., Suret, M. B., Bryant, C. M. L., Chapman, S. J. A. & Millar, L. (2009) A comparative analysis of the categorization of multidimensional stimuli: I. Unidimensional classification does not necessarily imply analytic processing; evidence from pigeons (Columba livia), squirrels (Sciurus carolinensis), and humans (Homo sapiens). *Journal of Comparative Psychology* 123(4):391–405. doi: 10.1037/a0016216. [JOvH]

Winner, E. (1997) *Gifted children*. Basic Books. [HG]

Woodley of Menie, M. A., Fernandes, H. B. F. & Hopkins, W. D. (2015) The more g-loaded, the more heritable, evolvable, and phenotypically variable: Homology with humans in chimpanzee cognitive abilities. *Intelligence* 50:159–63. Available at: http://doi.org/10.1016/j.intell.2015.04.002. [aJMB, MAWoM, RA]

Woodward, J. (2011) A philosopher looks at tool use and causal understanding. In: *Tool use and causal cognition*, ed. T. McCormack, C. Hoerl & S. Butterfill, pp. 18–50. Oxford University Press. doi: 10.1093/acprof:oso/9780199571154.003.0002. [IJ]

Wooldridge, D. E. (1968) *Mechanical man: The physical basis of intelligent life*. McGraw-Hill. [aJMB]

Yamamoto, M. E., Domeniconi, C. & Box, H. (2004) Sex differences in common marmosets (Callithrix jacchus) in response to an unfamiliar food task. *Primates* 45:249–54. [aJMB]

Yoerg, S. I. (2001) *Clever as a fox. Animal intelligence and what it can teach us about ourselves*. Bloomsbury. [aJMB]

Zullo, L. & Hochner, B. (2011) A new perspective on the organization of an invertebrate brain. *Communicative & Integrative Biology* 4:26–29. [KEH]